Effects of Computer-Based Simultaneous Listening and Reading on Second Language Vocabulary Acquisition

Philip George Goertzen

Ph.D.
The University of Edinburgh
1995

Acknowledgements

This work would not be complete without acknowledgement for the assistance I received in its conception and completion. For finding a manageable project, for inspiring discipline, and for deadlines, I thank my first supervisor, Eric Glendinning. Without his guidance and keen sense of how to get things done, this work would never have been finished. For his expertise in writing style and just the right mix of criticism and support, I would like to thank Tony Lynch, my second supervisor. For translating the questionnaire and for insight into Japanese students, I am grateful to Luli Ishikawa and Miki Inoue. Thanks are also due to Dan Robertson for his help with statistics and to Professor Alan Davies for his help with the multiple choice test and questionnaire design.

Paul and Heidi Hughes, Bill and Barb Daniel, Pam and David Hamilton, Andy and Fiona Drane, Chris Highcock, Gillian Donald and John McEwan have all been loyal and supporting friends throughout the duration of my time in Edinburgh.

For supporting us financially and in prayer I am thankful to Cornel and Katy Unruh and George and Kay Goertzen.

I would like to thank Rebecca and Amy for the much-needed motivation to finish this thesis. Finally, for inspiration and patience, for love, joy, and friendship, I thank Noni, my wife, to whom this thesis is dedicated.

On the heights along the way,
Where the paths meet,
Wisdom takes her stand.
(Prov. 8:2)

Declaration

This thesis has been composed entirely by myself.

Philip Goertzen

Abstract

This thesis evaluates the effects of computer-based, user-controlled simultaneous reading and listening on second language vocabulary acquisition. The experimental design consisted of two parts. First, 59 post-secondary school Japanese English students read a short story (approximately 1000 words) on the computer and then completed a multiple choice vocabulary test. The students were randomly divided into control and experimental groups. Both groups read the same story but the experimental group had the option of listening to the story, sentence by sentence, while they were reading. The aim of this part of the study was to determine if the post-test means of the reading-while-listening group would differ significantly from the post-test means of the reading-without-listening group.

In the second part of the study, 43 post-secondary school Japanese English students read the same story as above but instead of the multiple choice test, were given a 23 item questionnaire in Japanese. The first 20 items used a 5-point Likert scale to examine such issues as previous computer experience, enjoyment of the system, self-reported lexical and content comprehension, and assessment of difficulty of the text and the interface. The questionnaire also included 3 open-ended questions where students could comment on the materials.

A summary of the results is as follows. The vocabulary post-test results showed no significant difference between group means ($\alpha = .05$). The log files also showed a very low rate of listening to individual words but a comparatively high rate of sentence listening. Analysis of the log files showed no significant correlation

between word listening and post-test scores and only a weak positive correlation between the amount of sentence listening and post-test scores. Analysis of the questionnaire data revealed that: (1) students in the experimental group claimed to enjoy the experience significantly more than those in the control group; (2) the control group indicated significantly greater comprehension of content than the experimental group; (3) both groups enjoyed using the computer reading but did not prefer it to traditional media. Also, there was a strong correlation between previous experience and ratings of story content difficulty in the experimental group.

The results suggest that, for this population at least, simultaneous reading and listening does not have an immediate, positive effect on short-term lexical acquisition. Questionnaire results suggest that computer-based reading (without listening) is not disliked by students but that the addition of sound added to the enjoyment of the reading exercise for the population under examination. Relative to the control group, experimental group students tended to give themselves poor ratings on story content comprehension. Results from the open-ended questionnaire items suggest that despite the generally low level of previous computer experience, students have clear expectations of what computer-based materials should provide for the user. The implications are twofold: (1) the students expected an easy and intuitive user interface, and (2) the students favoured hypermedia materials, (i.e., text, sound, animation, video, and graphics) over computer-based, text-only materials.

List of Tables and Figures

Table of Contents

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Appendices

Introduction to the Thesis

Investigation into the development of computer-based language learning materials is at an exciting stage. The availability of inexpensive yet powerful machines and the simplification of development software has opened up unparalleled opportunities for the creation of stimulating language learning materials. The creation of materials has begun in earnest in many language teaching institutions, yet the effects of these new materials on language learning are relatively unknown. Many proponents of computer-based materials have argued that the materials are motivating, stimulating, and conducive to learning, but few have examined particular implementations of the technology in terms of specific pedagogic criteria such as comprehension and acquisition, enjoyment and ease-of-use. This thesis is intended to be a partial remedy for this problem.

In most modern computer-based materials, the machines are not doing something completely novel, they are instead, manipulating traditional ideas and methods, combining them in unique ways, and presenting them to students at great speed—the parts may not be novel, but the whole, the *implementation* of the materials most certainly is. In computer-based materials, we are seeing text, video, animation, narration, and photographs (none of them new in themselves) all in one place and all under the control of the student. This is not something that has been previously possible.

The particular implementation of materials under investigation here is the addition of sound to text. The specific pedagogic interests are vocabulary acquisition, and enjoyment, comprehension, and materials design improvement. The addition of sound to text was chosen because it represents one of the easiest ways in which non-expert designers (i.e., interested language teachers) can develop multimedia computer-based materials. It was also felt that this kind of research might suggest what kinds of materials could be created to encourage greater vocabulary acquisition, and greater enjoyment of reading.

The desire of this researcher is that the thesis should be practical. Broadly speaking, the primary aims include the following: It should suggest one or two ways in which English can better be taught and learned; it should use the insight gained from theoretical research to improve computer-based materials; and it should inspire further investigation and development into the use of computers for language teaching. It is hoped that these aims will be realised in the six chapters that follow.

1.1 Overall Organisation of the Thesis

In addition to this introduction, the thesis consists of six chapters. Chapters 1 and 2 review the literature on computer assisted language learning and second language vocabulary acquisition, respectively. Chapter 3 describes the design of the experimental procedure and includes the experimental method. Chapter 4 delineates the results of the statistical procedures used to test the null hypotheses. Chapter 5 discusses these results, and Chapter 6 provides a summary and conclusion for the thesis. The bibliography appears after Chapter 6. Appendices to the main text appear after the bibliography.

1.2 Chapter Organisation and Contents

Each chapter includes its own table of contents, introduction, and summary. Where appropriate, key terms will be defined after each introduction followed by the main arguments of the chapter. The following six sections give a more specific overview of each chapter.

1.2.1 Chapter 1

Chapter 1 is primarily concerned with the use of computer-based materials in education and second language teaching. The chapter examines the development of computer assisted instruction over the past four or five decades. Specific attention is paid to the literature supporting the use of computers in education, the strengths of the argumentation in this literature, and the changes that general educational use has brought to computers in language learning. Towards the end of Chapter 1 the emphasis is twofold: (1) the use of hypermedia materials specifically (hypermedia is defined in Chapter 1), and (2) the gaps in our understanding of hypermedia use in foreign language teaching.

1.2.2 Chapter 2

Chapter 2 reviews the literature of second language vocabulary acquisition. Like Chapter 1, the initial discussion adopts a historical perspective. The nature and role of vocabulary in language teaching over the past four or five decades are discussed. This includes an examination of the importance of vocabulary (relative to other theoretical interests) over the years and suggests reasons why vocabulary is important. Moving from this broad discussion, the chapter progresses to a brief review of the first language mental lexicon, the difficulty of learning words, and the nature of the second language mental lexicon. The emphasis of the chapter then shifts to incidental vocabulary acquisition. In particular, various research studies are presented and the suggestion that simultaneous reading and listening could be a good source of incidental vocabulary acquisition is posited. The chapter ends by bringing together chapters 1 and 2 with a discussion of the appropriateness of a hypermedia program for reading while listening.

1.2.3 Chapter 3

Chapter 3 explains the experimental method used to examine the effects of reading while listening on vocabulary acquisition, enjoyment, comprehension, and various student attitudes. The chapter describes the text to be used, its selection procedure, the way sound was attached to the story and the way students were able to elicit the sound. The experimental design consists of two parts. The first part uses a control group/experimental group post-test design. Students in both groups read the same text but the experimental group students have the option of listening to the text while they read. The groups are post-tested using a multiple choice vocabulary test. The items of this test were taken from the story and the correct answer was the meaning of the word as it appeared in the context of the story. Students' actions are recorded by the computer for later assessment. The null hypotheses in this section are that there will be no significant difference between the post-test means of the two groups and that there will be no significant correlation between the amount of listening a student did (as measured by the log system) and their post-test scores.

The second part of the experiment uses a control and experimental group who read the same materials as the groups above. However, instead of a multiple choice post-test, the students complete a questionnaire after reading the text. Most of the questionnaire items are statements with which students are asked to agree or disagree on a five-point Likert scale (including one 'don't know' category). However, comprehension questions required students to estimate their understanding of overall vocabulary and story content in percent, on a five-point scale. Similarly, questions of previous computer experience ask students to estimate the frequency of previous computer use on a five-point scale. The questionnaire also includes three open-ended questions about the suitability of the computer exercise and suggestions for improvement.

There are many hypotheses associated with this questionnaire, but the primary null hypotheses are that there will be no significant difference in the distribution of responses between groups to the questions of (1) vocabulary and plot difficulty, (2) vocabulary and plot comprehension, and (3) enjoyment of the reading exercise. All null hypotheses and their alternative hypotheses are listed and assigned numbers for reference in Chapter 4.

1.2.4 Chapter 4

Chapter 4 provides an analysis of the data collected for Chapter 3. In the first part of this chapter, the characteristics of the experimental sample are discussed. Following that, each null hypothesis is examined in turn, starting with the post-test group and then the questionnaire group. Null hypotheses are rejected on the basis of the statistical analyses. At the end of this chapter, significant results not specifically anticipated in Chapter 3 are presented. Included in this is an analysis of the questionnaire group's previous experience, their overall consideration of the difficulty of the experiment, and their overall opinion of the computer's usefulness as a reading tool.

1.2.5 Chapter 5

The first part of Chapter 5 discusses the data from the previous chapter. It includes first a discussion of the post-test results, then an examination of the questionnaire results. The qualitative data from the open-ended section of the questionnaire is used to suggest explanations for various results. Then, later in the chapter, it is used to suggest improvements that might be made to the computer-based activity for classroom use. Finally, the qualitative data are used to provide a basis for determining student expectations of hypermedia materials—in the hopes of making generalisations about what constitutes good hypermedia documents.

1.2.6 Chapter 6

The first part of Chapter 6 summarises the previous five chapters. It then draws conclusions based on the data and discussion. The last part of Chapter 6 discusses areas where the present research might be continued and suggests related areas where future research might be carried out.

Table of Contents: Chapter 1

Chapter One
Computers, Hypermedia, and Research Issues

Introduction

In this chapter we will explore the field of computer-assisted language learning (CALL) from two perspectives: (1) theoretical and practical developments in an historical context and (2) research requirements in the modern context. Section 1 will provide some working definitions of key computer-related terms. Section 2 will examine the history of computer assisted instruction (CAI) over the past four decades. Specific attention will be given to the theoretical context in which CAI developed and the corresponding theoretical climate that accompanied CALL. The section will progress from a general discussion of computers in education to a focused examination of hypertext and hypermedia. Section 3 will discuss various theories that researchers argue support learning with hypertext and hypermedia materials. Section 4 will discuss the need for more research in hypermedia, suggesting several reasons why there is a shortfall in experimental literature, a rationale for conducting more research, and several key areas where research could begin.

1.1 Working Definitions

1.1.1 Hypertext

Hypertext is the name given to a computer-based document incorporating electronic links between chunks of text on the screen and other chunks of text somewhere else in the computer file. In the design phase of hypertext materials, links are established by an 'author' who creates the 'document'. When the document is finished, the user can activate the links by clicking with the mouse on specially highlighted words (often bold, italics, or coloured text).The software then jumps to the new chunk of text and displays it on the screen. As a popular example, we could consider Internet 'browsers' like *Netscape* and *Mosaic* and the *Windows* 'help' system to be typical hypertext systems.

1.1.2 Hypermedia and Multimedia

Hypermedia software is exemplified by any of a large number of computer authoring programs designed to control and enhance multimedia devices such as video, sound, animation, hypertext, text, and graphics. Users are given a high degree of control over these devices and experience many possible stimuli—textual, graphical, and aural. Hypermedia is available over a wide range of computer platforms, including desktop PCs, Apple Macs, Unix systems, and others. (See Appendix 1.1 for an overview of various hypermedia software packages). The terms *hypermedia* and *multimedia* are often used interchangeably; we do not take issue with this. However, before the advent of the IBM PC, multimedia generally meant slides (or video), text, and music, with a limited coordination of these media. There was no implication that the materials were computer-based (although the mechanism controlling them was usually a simple dedicated computer system). Some people still understand multimedia in this way. For this reason, despite the arguably inelegant nuances of the prefix 'hyper', *hypermedia* will be used exclusively in this thesis. Ingraham, Chanier and Emery (1994: 107) have used the following definition in reference to multimedia, but we will use it here as a working definition of hypermedia:

> ...a computing resource capable of displaying textual data, graphical data (still or animated), photographic data (still or moving), and of recording and playing back sound data. These various types may be displayed simultaneously or sequentially depending on the needs of the resource....Most large-scale systems will provide multiple data *Windows* and...these *Windows* should be capable of displaying different data types.

When using the term *hypertext* one is usually referring to a type of software that makes use of a specific type of human-computer interaction (searching and browsing through 'jumping'). Hypermedia, on the other hand, may very well include hypertext as a means of controlling and navigating through various computer-based media. The term hypermedia stresses the materials themselves, while the emphasis on hypertext is on the type of interaction.

1.1.3 Interface

There are two nuances to this word that are relevant to this study. First, the word *interface* is used in a general sense to describe the type of operating system that a particular computer uses. A graphical user interface (GUI) uses pull-down menus or metaphors to give the user a choice of possible actions (e.g., erasing or copying files) while a command driven interface (CDI) makes use of typed commands (which the user must learn beforehand) to effect the same action ('metaphor' is used in the computer field to mean a graphical representation of a familiar physical object). Typical examples include pictures of file folders where files are stored, pictures of rubbish bins where files are erased, or pictures of a check mark to indicate a spelling checker). The second meaning of *interface* is slightly more specific. It is the physical means of interacting with a particular program within the bounds of the specific operating system. In this sense, when we refer to the *physical interface* or *interface device*, we mean an input device such as a mouse or keyboard. This is contrasted with the first kind of interface—often referred to as a *screen* or *graphical interface*—which is a reference to the way the keyboard or mouse is represented on-screen and the metaphor that is used to indicate to the user that something should be done.

1.2 History of Computer Assisted Instruction

1.2.1 Programmed Learning

In the 1950s and 1960s, American educators were already beginning to envision 'teaching machines'. Richmond (1965: 34) wrote:

> What was needed, more than anything else, was a drastic revision of the conventional text-book format, a method presenting information and ideas so as to ensure continuous, unremitting participation on the part of the readers.

The use of computers for the control of the user's participation in learning was initially a part of *programmed learning*; the proponents of programmed learning in the 1960s were optimistic. Oettinger (1969: 4) wrote,

> In the first tableau of our vision, we see that it is technically possible for sound, pictures, and even objects stored at appropriate

> centers to be available with the greatest of ease and negligible cost at innumerable local points of access, first perhaps in schools, libraries, or factories and only a little later, as seen from our visionary perspective, in every home.

Oettinger's vision, while technically possible in the 1960s, was practically and economically unfeasible. The details of his vision, however, are not only possible today, but are already in use. The four areas where Oettinger envisioned unlimited access (access to catalogues of the great libraries; catalogues of film libraries; experimental teaching programs; and tools to manipulate resources like dictionaries, thesauruses, and mathematical programs), are all readily available in the 1990s (in some form) by way of the Internet and other Internet-based resources (e.g., Compuserve, GOPHER, etc.).

1.2.2 Difficulties

The implementation of computers in educational environments brought with them a new set of difficulties. These included a misunderstanding of technology, fear of computers, and difficulty of operation. Fear and misunderstanding arose for several reasons. When computers were installed in North American schools, advocates of the computer 'establishment' (manufacturers, academia, and programmers) touted computers as 'thinking machines' and 'human replacements'. The labour force in the manufacturing industry felt rightly threatened by computers and computer-aided machines.

Consider briefly, for example, this sample of titles and overviews of articles in the popular American business magazine *Business Week* from the 1970s and 1980s (source is ABI/INFORM Full Text Database):

'AUW Fears Automation Again' (1977)

- United Autoworkers fear increased automation will lead to greater unemployment.

'A Technological Strike at ABC' (1979)

- Network television cameramen fear job loss as automated cameras begin to replace the traditional three-man crews.

'How to Conquer Fear of Computers' (1982)

- Managers in business fear computers because of the risk of looking foolish.

In the more scholarly business journals the concern was also evident. Articles include:

'Effects of Technological Change on Trade-Unions' (1975)
Production Journal

- Technological change presents extreme danger to employment for trade union members.

'New Office Technology and Employment' (1980)
International Labour Review

- Low-level jobs will soon be lost to computers.

Articles like these suggest that the attitude towards computers was initially one of fear and/or mistrust. Computers *were* replacing people on assembly lines and white collar employment was suffering also. It is not surprising, therefore, that there was resistance to the implementation of computers in schools and universities. During the 'everything is possible' 1950s in America (and the two decades to follow), even teachers thought their jobs were at risk to computers.

Threatened job loss, however, was a minor impediment to technological implementation compared with the level of expertise required to use early computers. Richmond (1965: 35) wrote that:

> For some teachers, the prospects opened up by these developments [in machine-based learning] may at first seem startling, not to say disturbing. Many of the claims made for the new 'method' strike

> them as being extravagant. To make matters worse, much of the current literature on the subject seems to them to be couched in a technical jargon which is unfamiliar, or even incomprehensible.

It would take a long time until the parts and vocabulary of the computer were as familiar to the average person as were, for example, the parts and vocabulary of an automobile. It has been a common argument over the past twenty years or so that the 'next' generation, having grown up with computers, would be more likely to utilise technology and would be at ease with technological change. Twenty-five years after Richmond's above comments, however, Underwood and Underwood (1990:17) investigating the technological skills of new teachers write that,

> Not only did the under- and non-users express doubts about their own abilities to use the technology, but they were also unsure as to what to do with it in the classroom. Hence, many took the path of least resistance either by using the machine for drill-and-practice or not using it at all.

Experience and common knowledge support the notion that computers are *still* poorly understood and under-used by the majority of teachers and students. Why is it that so little has changed in twenty-five years? There are at least four possibilities: (1) those who grew up with computers never learned how they worked and never learned skills to develop/manipulate computers for their own use; (2) the majority of educators are still of the 'last generation' and the technically-literate generation is still to come (cf. Higgins and Johns 1984); (3) because the field changes so quickly, and systems/software are still difficult to use, teachers find it impossible to stay current with the technology; or (4) most teachers (of any generation) know how to use computers for some purposes but do not feel the need to learn how to teach with them.

In addition to these four possibilities, informal discussions with other language teachers have revealed several points suggesting why computers are not used. These are:

1. computers produce anxiety (as it will be defined in section 1.4.3.1);
2. teachers do not know what it is possible to create;
3. teachers associate existing computer-based materials with out-dated theories and claim (sometimes rightly so) that computer-based materials could be easily replicated with paper and pencil activities (cf. Higgins and Johns 1984).

Of these three, the first two can be overcome with experience and awareness-raising. The third point, however, deserves a more detailed discussion—the beginnings of which should include an historical look at CAI.

1.2.3 CAI in a Theoretical Context

Early technology-based instruction was well suited to the prevalent educational theories of the 1960s—however unpopular these theories might be now. The capabilities of computers to manage text and numbers and present students with simple problems and simple rewards, suited theories of stimulus/response and other behaviourist ideas (cf. Skinner 1954). However, machines were more appropriate where behaviourist methods like *operant conditioning* are successful, i.e., at the very low levels where "the learner gets most of the answers right and receives an unbroken series of *reinforcing stimuli*". (Richmond 1965: 43). Early computer programs asked questions and expected users to either choose the correct answer from a list or enter what they thought was the right answer. Early systems did (and modern systems still do) have difficulty accommodating questions with a range of correct answers. In the simplistic learning context of behaviourism, however, these types of system were not controversial. Proponents of programmed learning accurately argued that programmed learning was most appropriate in the context of teaching 'subsidiary skills', including (but not restricted to) grammatical inconsistencies, vocabulary, synthesis and description of narrative writing, and teaching stylistic appreciation (Hilton 1974).

Behaviourism and stimulus-response theory eventually gave way to educational theories with more explanatory power. In cognitive psychology, studies into human memory became the dominant interest of experimental cognitive psychology (Carroll and Mack 1984). In the 1970s, the psychological literature was primarily concerned with the notions of long and short term memory (e.g., Norman and Rumelhart 1975). Indeed, twenty years later, studies in memory still account for a great deal of the investigative effort in cognitive psychology.

1.2.4 Changes in CAI

Computer-based learning in general attempted to join the retreat (or advance) from behaviourism to more comprehensive learning theories. *Linear Learning* (as stimulus/response learning is sometimes called) was adapted and *branching* programs were developed (Richmond 1965). Learners who used branching programs not only read materials and answered questions, but also were given choices for answers that then referred them to other pages/books for the explanations to the answers (and both right and wrong answers would have explanations). The learner would continue from the place to which s/he was referred, jumping here and there throughout the materials. Early computers and systems were ideally suited to this type of learning—if for no other reason than they could 'turn' the pages for the learner and that more information could be stored than in printed volumes. However, linear programs that implemented branching usually were only a specialised departure from the norm—and they were already being seen as less than adequate (Richmond 1965).

It was the development of branching programs, however, that introduced the concept of machine/user *interactivity*. The control of the learning was shared by computer and user; the degree of control dictating the level of interactivity (see section 1.4.3.3 below). Interactivity became a kind of buzzword and eventually became the goal of most educationally minded materials developers. However, the factors determining the interactivity of programs were not investigated until the introduction of the arcade game in the latter part of the 1970s. Observations of children playing computer games inspired many to examine the various aspects of a game that produced motivation. For example, enthusiasts like Malone (1984) observed that arcade games could hold the attention of users over extended time periods (Malone also made the valuable distinction in his work between *toy* and *tool*—A toy is easy to learn but challenging while the use of a tool is simply a means of accomplishing a goal and should remain almost invisible to the user). Observations were also made at this time that good teaching materials should: (1) be easy to learn but difficult to master; (2) be surprising but not incomprehensible; and (3) make use of sound and graphics in addition to text (Malone 1984).

1.2.5 Changes in CALL

On the basis of research done with arcade games and other computer-based 'toys' a rationale for teaching materials began to develop in the 1980s for CAI. Unfortunately, it is difficult to trace a parallel development within foreign language teaching. The post-behaviourist years of the 1970s and 1980s saw a growth in the number of theories of second language learning and acquisition. For example, in the later part of the 1970s, under the influence of Corder and later Selinker, the concept of interlanguage took root and began to dominate the thinking among researchers (James 1981). At nearly the same time, interest grew in morpheme acquisition order, followed shortly by Krashen's monitor model and recently into an eclectic theory (that emphasises learner differences and styles while de-emphasising the apparent clarity of the learning/acquisition dichotomy) exemplified by researchers like Ellis (1992 and see section 1.2.7).

1.2.6 Post-Behaviourism in 1980s CALL

CALL development in the 1980s was characterised by developments in hardware that made possible programs that could expand on programmed learning while paralleling other advances in language teaching theory. Stevens (1992: 11), writes that in this and the previous decade, CALL is "...shaking off the influence of the early behaviourists...". In the post programmed learning era, he argues, software is characterised by humanistic objectives rather than the discrete (and non-communicative) goals found in programmed learning and behaviourism (Stevens 1992). The theoretical outcomes of this paradigm shift were that users and developers began to expect software to be easy to use; provide choices and allow for different approaches to learning; and most of all, that it permit learners to learn for themselves. The practical manifestations of humanistic objectives were materials that provided choice to the learner and teacher (e.g, authoring software like *Storyboard*) , a focus on learning through investigation rather than teaching through drill and practice (e.g., hypertextual glosses to text), and an emphasis on exploratory interaction with realistic materials (e.g., classroom-based concordancers using authentic corpora). Eventually, the growth of computer-based problem-solving games and hypermedia contributed to the new sensitivity to the learner while accommodating non-behaviourist, humanistic theories of language acquisition (Stevens 1992).

1.2.6.1 Authentic Text, Concordancers, and Learning through Context

When looking at a specific time period in the ever-changing field of CALL, it is often beneficial to look at what advantages the computer provided in that time and how or to what degree that advantage was utilised by CALL developers. One should evaluate CALL with this in mind (rather than the much more common practice of evaluating CALL against some unrealistic dream of what should be possible). In the 1980s, CALL developers and enthusiasts for the most part had only efficient text-handling at their disposal. Concordancing software was an example of using this capability well. Combined with authentic text, concordancers exemplified the clever (if focused) use of hard/software within a sound theoretical framework.

The push for authentic texts was supported by the work of researchers such as Johns (1989 cited in Stevens 1992) and Stevens (1992) who argued fairly successfully that simplified texts, contrived texts, or 'doctored' texts (i.e., texts where, for research purposes, vocabulary items or grammatical patterns were made to appear in an unnatural frequency) bore little resemblance to texts that might be used in the classroom. The data from experiments that made use of inauthentic texts was arguably less valid than data derived from authentic texts.

In the 1980s, the development of concordancing programs for desktop computers brought the data from authentic texts (in terms of collocational information useful for vocabulary learning and rule formation) to the average user, i.e., teachers and students. Concordance programs were traditionally associated with applied linguistic research and mainframe computing (Stevens 1991b). This perception quickly changed, however, with the work of researchers such as Johns (1986) and later Stevens (1991b) who argued that advanced students could utilise concordancing in self-study. The concordancer provides immediate access to vast amounts of lexical and grammatical information. Students' research of variable English rules and collocational information about words could be enhanced as they examined texts with the software.

It is the fact that concordancers usually make use of naturally occurring texts (textbooks, newspapers, magazines, literature, and popular novels) and the fact the information gleaned by students through concordancers is valid (i.e., relevant to 'real English') that makes the use concordancers so appealing. Students who use concordancers can trust the information that they have about vocabulary because they have witnessed the use of the word in context (usually many contexts). Similarly, they can trust the rules that they establish for themselves because their knowledge comes from experience with the language rather than from rules acquired from textbooks or teachers. Furthermore, as Johns (1986: 159) notes, concordancers provide an intermediate position between inauthentic textbook texts and "the potentially confusing but far richer and more revealing 'full flood' of authentic communication." For CALL and language learning in general in the 1980s, the use the concordancer accomplished three things: (1) it facilitated the growing desire among second language researchers for students to learn through authentic texts; (2) it facilitated the growing interest in second language vocabulary; and (3) it facilitated the potential of students to work independently.

1.2.6.2 Choices

Choice is characteristic of hypermedia-based language materials. However, the concern for learner choice is evident in the literature well before the advent of mainstream hypertext and hypermedia authoring programs. Legenhausen and Wolff (1990: 1), for example write that the "collective wisdom of the profession" is that there are different learner personalities and types, with different learning styles and motivational orientations. There is also a wide variety of language features that are acquired in different ways, and there are different learning settings which will affect the acquisition process. Software, they argue, should permit and encourage learners to use a variety of strategies; the strength of that software depends on its ability provide the choice of strategies to the learner.

1.2.6.3 The Progression to Hypermedia

The excitement with hypermedia in language learning, its concern with choice, freedom, interactivity, and communicative authenticity should be seen as a result of developments in the previous decade—both theoretically in making the computer accommodate contemporary notions of language acquisition, and practically in software that provided as much choice, freedom, exploratory learning, and authenticity as was physically possible. While the first authoring tools could be customised by teachers for students in what was formerly a completely designer-controlled enterprise, they were still far less functional than might be hoped for in the theoretical milieu of the time. The idea of teacher-designed materials was good, but the end result (i.e., what the student saw) was still linear (students worked from beginning to end) and uni-dimensional (students were limited to textual interaction only). However, the development of graphical hypermedia authoring systems early in this decade (which only really became possible with the availability of more powerful computers) is a further step closer to flexible (the teacher can determine the content), communicative (the student works in a computer-generated context that can, to a degree, mimic real-world contexts that require listening, reading, and writing), and non-structure-based (the computer can deal with more than just drills) computer materials. Most would agree that these characteristics are prerequisites for modern language teaching.

1.2.7 The Computer's Strengths and Weaknesses

1.2.7.1 Traditional Strengths

The potential of present day authoring tools might better be understood with the computer's strengths and weakness in mind. The computer is well designed for some tasks, poorly designed for others. The computer, by definition, *computes*. This means that it calculates mathematical formulas and arithmetical computations (often colloquialised as 'number crunching'). From earliest times, computers were designed, constructed, and implemented, to replace the task of manual addition and subtraction, multiplication and division. The computer not only does these things for which it was designed, but it also does them millions of times faster than the human mind and will do them, over and over, tirelessly, for as long as it is instructed to do so, without making mistakes. What does this great mathematical ability have to do with language learning? On the surface, and in the early stages of computer development, there were few obvious applications.

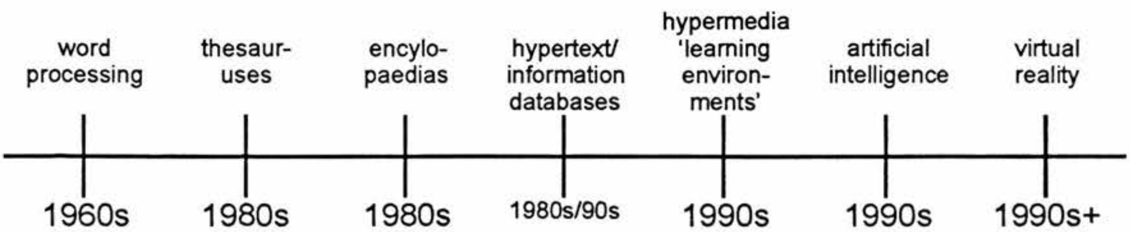
All written languages, however, can be transcribed in terms of numbers. Indeed, the computer only understands its own programming language after it has been translated into numerical form. Since the computer recognises words as numbers, and is very good at dealing with numerical data, it can very easily manipulate, search for, and retrieve text. This is not to say that computers can manipulate language (in its grammatical and/or cultural context)—it can only manipulate text. Thus, text-intensive linguistic tasks such as searches for dictionary items, related ideas (through a database), or graphics related to ideas, can be done easily and with great speed. The advantages of computer use in this context are obvious, and as has been said, relatively uncontroversial. More recently, the ability of computers to digitise sound, photographs, and videos (once again, turning apparently non-numeric data into numeric form), has allowed the potential development of materials that not only speed up traditional tasks but create new tasks—tasks possible only through computer-based activities.

1.2.7.2 Modern Strengths

The face validity of computer-based materials in the post-communicative context requires that there be graphical and aural/oral interaction as well as data on culture and context; it is likely that students expect this and there is no reason why designers can not provide it. The computer is becoming more and more capable of offering a degree of this kind of interaction. However, the discussion in the literature of these possibilities is at an early stage.

The underlying questions of any thesis aimed at testing the effects of computers on learning must still be: what, if anything do computers have to offer the language teaching profession, and which aspects of computer use offer the simplest, most effective opportunities for the interested teacher?

Within CAI, there is a range of software that might assist the study and learning of language/s. On one end of this range is the purely functional word processor, at the other end artificial intelligence and virtual reality. Closely related to word processors are tools such as concordances, spell checkers, dictionaries, thesauruses; while nearing the right end of the scale, we find such things as talking dictionaries, talking books, games and software designed to create 'learning environments', hypertext, hypermedia etc.. The following table summarises the situation:



All points on this scale carry a degree of controversy. The difficulties arise as we move rightwards on the scale. This is where the computer begins to replace (or improve) traditional teaching tools and practices. Indeed, some would have the extreme end of this scale replacing the most traditional of all 'tools'—the teacher. It is at this end of the scale, however, where the interesting questions need to be answered, where the effects need to measured, and the responses of students need to be considered. It is the possibility of combining any or all of the above that makes hypermedia a viable learning and research tool in the framework of current language learning theory.

Obviously, the least controversial (and best understood) uses of computer are also the ones that have been with us the longest. In the beginning stages of computer development, computers were only capable of acting either as simple tutors (as discussed above) or as tools (usually offering improvements in access speed on traditional tools such as dictionaries and data collection). As hardware and software develop, they might be expected to be more interactive, more anticipatory of users' actions, and more likely to make relevant suggestions within an interactive context.

The researcher must therefore find a place on this scale where the technology is mature enough to be make development and research possible, and new enough to make the answers to the research question relevant. Furthermore, the materials developed should fit within the current theoretical context.

1.2.8 Hypermedia within a Modern Theoretical Framework

Current theorists and practitioners embrace some or all of the following arguments to a certain degree (adapted from Ellis 1992):

1. Learners vary from each other both individually and in language groups;
2. Learners possess a number of styles;
3. Learner styles may change as the perception of a task changes;
4. Learners' production varies in accuracy depending on the task;
5. Learners have different communication needs at different times;
6. Unless students are learning a language for a specific reason such as taking an examination, the primary purpose of instruction is for students to improve their communication skills (communication can be verbal or written);
7. Learners should be permitted to partake in 'unfocussed' 'authentic' communication that might take the form of problem solving; filling gaps in knowledge; negotiation with interlocutor to establish meaning; and the control of language learning resources;
8. Distinctions can still be made between acquisition and learning.

This is by no means an exhaustive summary of current language learning theory. It is given, instead, as a framework within which modern materials (both teacher and computer-based) should be expected to work. Hypermedia *is* certainly more capable of working within this framework than traditional text-only computer-based materials. The degree to which it is successful depends primarily on the skill and creativity of the developer and the persuasiveness of CALL researchers. There are, of course, several caveats that must be considered before we even begin to argue the merits of hypermedia as a language teaching tool:

1. In terms of communication, computers are better at assisting the written compositional needs of students than they are at assisting their oral/aural needs;
2. Computer can not yet adequately respond to a user's speech;
3. Students can not interact to the degree that a teacher can (see 1.4.3.2 below);
4. Computers are relatively expensive and require expertise both to teach with and to learn with.

With these four conditions in mind, we will address the question of hypermedia's suitability first from a historical perspective and then from a research perspective.

1.2.9 Hypertext History: Mechanical

The general history of computers in education and language learning has been discussed but because hypertext was conceptualised from several traceable theoretical perspectives, its history is worth exploring here. Despite the fact that the program described later in Chapter 3 does not use hypertext *per se* (links are to sound rather than to text) it is driven by a hypertext authoring program and future expansion of a program on this model would almost certainly make use of some hypertext features.

At the birth of the computer era in the 1940s, scientists and researchers dreamed of a means of linking the vast amount of data that was created in the post-war years. Huge research projects such as the development of jet aeroplanes and research in nuclear fission resulted in untold volumes of paper and massive file systems. The computer technology which is commonplace today was only a light in the eyes of the richest American researchers. One of those who dreamed of a better way of organising information was Vannevar Bush, administrator of the Manhattan Project and President Roosevelt's chief scientific advisor (Galletly, Butcher and Daryanani, 1992). Bush foresaw an intricate and sophisticated machine called a *Memex*. Hardman (1990: 2) quotes Bush's 1945 seminal work:

> It [the Memex] affords an immediate step, however, to associative indexing, the basic idea of which is provision whereby any item may be caused at will to select immediately and automatically another.

The Memex would use microfiche linked by mechanical levers and be accessible from a central information desk (Woodhead 1990). Anyone with access to the machine would have access to all the information in the Memex and would have a greatly simplified research process. Bush's peers considered him eccentric. Alas, a mechanical device was simply impractical.

The 1960s, however, brought hope to dreamers such as Bush. Large companies and government bureaucracies began to use mainframe computers to keep track of clients, charge taxes, and automate labour intensive information processing. Another visionary, Douglas Engelbart of Stanford University, making use of a mainframe, conceived and created an on-line system of information retrieval. The system was later developed by McDonnell-Douglas into a simple but functional database system (Woodhead 1990). Hierarchically structured, expressions entered by the user could retrieve any text in any file within the system (Hardman 1990). McDonnell-Douglas was one of the first large companies to use computers for this kind of information technology.

The term *hypertext* was coined in 1965 by Theodore (Ted) Nelson (Jacobs 1992). Nelson had mainframe computers at his disposal and could consequently anticipate a time when all information would be stored on electronic media. The purpose of Nelson's conception was to make knowledge and information accessible across a variety of academic disciplines. Indeed, Nelson foresaw a time when boundaries between traditionally disparate fields (e.g., literature and biology) would be removed because the electronic links between topics would so increase the speed of research and investigation that 'everything' would be available to everyone (Woodhead 1990 and Hardman 1990). Nelson's vision was more practical than Bush's (computers being a realistic means of accomplishing the task), yet more abstract than Englebart's (Nelson's dream of removing subject boundaries was at times grandiose and implausible). Nevertheless, it is this model of accessible knowledge across disciplines that continues to influence expectations of hypermedia software designers.

In the 1970s, technology began to catch up with the dreaming. Randall Trigg of Xerox Laboratories in Palo Alto, California, developed a software application called *Notepads. Notepads* was designed to take the place of index cards, the tool traditionally used by researchers to organise study notes according to topics (Woodhead 1990). Like its predecessors, *Notepads* ran on mainframes. It was this program that the United States Navy modified to organise their information and later became known as the Knowledge Management System (KMS).

In the early 1980s, developments in computer processors brought computers to the desktops of offices and homes in North America. The success of IBM's original PC motivated other companies to develop systems that would be easy to use and 'dummy proof'. As far as hypertext is concerned, the development that proved most significant was the introduction of Apple's Macintosh computer and the subsequent introduction of HyperCard in 1987. HyperCard used the index card metaphor introduced with *Notepads* and added a powerful programming language (called *Hypertalk*) that the popular press (e.g., Tessler 1995) claimed almost anyone could learn in a short period of time (compared with traditional programming languages). HyperCard was also 'bundled' free with every Macintosh sold in the United States (and many other parts of the world), leading to its widespread acceptance as the industry standard hypertext application (Woodhead 1990).

Hypertext applications in the IBM world developed slightly more slowly than in Apple's. In North America, the educational computer of choice was the Apple (though at first, not a Macintosh). When the Mac arrived students found it easy to use and Apple tended to cater to the educational market (IBM PCs being perceived as 'serious' business machines, suited only to number crunching and corporate database applications). Recently, however, with the general acceptance of *Windows,* software developers have started to think of the PC as a potentially powerful educational tool. One of the first programmers to contribute to hypertext for IBM/compatible machines was Peter Brown at the University of Kent (Woodhead 1990). Brown called his program *Guide* and later sold it to Office Workstations Limited

International (OWL) who have subsequently developed versions for *Windows*, Unix, and Apple Macintoshes.

1.2.10 Hypertext History: Conceptual

Enthusiastic hypermedia proponents support their arguments by attributing hypertext concepts to philosophical father models. McKnight, Dillon and Richardson (1991) write that notions of the interconnectedness of information can be found in the Talmud, and in the works of Aristotle. The point, however, is not that great minds conceived hypertext, but that humans tend to organise and recall things in certain ways—sometimes topically (encyclopaedias), sometimes alphabetically (dictionaries), sometimes apparently at random (personal 'filing systems'). In any case the idea of non-linear information processing is not new. One does not, for instance, read the entire Bible when looking for incidences of a particular word—one simply refers to a biblical concordance which cites all the occurrences of the word in question. Indeed, there are many written documents that make use of non-linear processes (Williams 1992 in Butler 1992). Among these are: newspapers (often read topically), dictionaries (read alphabetically in search of specific information), and instruction manuals (usually read with reference to an index). It is the speed and ease with which large documents can be linked and processed that create much of the excitement surrounding hypertext as an educational tool.

1.3 Theoretical Concepts in Non-Linear Learning

Proponents of hypertext have argued its educational benefits from a variety of viewpoints. McKnight *et al.* (1991) view the role of hypertext in education from three perspectives. The first is that reading in hypertext is somehow *a more natural process* than reading linearly. The second is that hypertext is an *augmentation environment*, and the third is that hypertext is simply a *storage and access mechanism*. Researchers such as McKnight *et al.* admit that we know very little about the cognitive processes of learning through hypertext. Nevertheless, several claims that fit into one or more of these perspectives should be discussed in terms of language learning.

1.3.1 Exploratory Learning

One such claim is that hypertext can be part of an 'exploratory learning system' (Hammond and Allinson 1989). Exploratory learning means simply that the students learn through personal choices based on a number of criteria including interest, perception of need, motivation, and appeal. This is not to say that learners are without guidance nor that the learning process is completely 'exploratory'. On the contrary, the teacher, to a degree, controls and feeds students interests (through planning and enthusiasm), creating needs (through tasks and assignments), and designing lessons that are both motivating and appealing. Hypertext documents, if carefully constructed, offer the student a large degree of freedom to explore and thus fulfil the criteria of exploratory learning.

Human beings often think tangentially and their curiosity is often sparked by various items that they come across while reading. It is thought by some that since hypertext documents facilitate inquisitive exploration and supply a means of querying certain words, phrases, or ideas, hypertext reading is a natural (by *natural* we mean that there are no external constraints in readers' inquisitiveness) process that should increase understanding and interest in the reader. Unfortunately, because the field is new and the body of literature still in the formative stages, it is difficult to assess these claims in terms of verifiable experiments. Thus the notion that hypertext reading is more natural remains an interesting but uncorroborated hypothesis.

1.3.2 Semantic Webs and Nets

Jonassen (1990) takes a slightly different approach to the processes involved in learning through hypertext. He argues that learning involves associating new items within frameworks of existing knowledge. Words, phrases, sentences, and even ideas are known only in terms of their relationship to existing knowledge or understanding. This kind of argument emphasises the semantic relationships between known and unknown rather than treating new material as discrete or independent points. With regard to hypertext, Jonassen believes that the system of linking words and phrases or even graphic images creates in the user's mind a 'semantic net or web' (Jonassen 1990: 115), facilitating the process of learning new information and enhancing the understanding of relationships between ideas.

1.3.3 Cognitive Flexibility Theory

The theory of cognitive flexibility was developed by Spiro and Jehng (1990) to provide a theoretical justification for teaching with hypertext to advanced learners. Spiro and Jehng argue that as content becomes increasingly more complex, it becomes more and more difficult to represent through a linear document. In fact, subject specialists often simplify their lectures to beginner students telling them only 'half the truth' in order not to overwhelm students with the vastness and complexity of the topic. Spiro and Jehng argue that linear documents are fine for subjects that are naturally well-structured and fairly simple. Linear documents are fine also for beginners—presenting a simplified subject at the beginning of a course is a common teaching strategy. Advanced learners, however, can benefit from a medium that reflects not only the complexity of the subject matter, but the interconnectedness of ideas, hypotheses, rules, and meaning. Spiro and Jehng feel that hypertext is such a medium. They use hypermedia to demonstrate the recurrence of themes in literary and cinematic works. The students learn that literary criticism is not straightforward and may be approached from a variety of perspectives.

Cognitive Flexibility Theory provides an interesting perspective for language learners. Complex relationships like those that exist between idiomatic phrases, their use, and their meanings, are difficult to teach explicitly to students but hypertext links might simplify the mystery. Words with apparently inexplicable spellings (e.g., knight, slough, etc.) seem inconsistent until links with related words or etymological descriptions are attached. Sociological and cultural descriptions with situational examples are difficult to provide but might be represented with links to graphic files, animated sequences, or video sequences (e.g., extracts from documentaries or newsclips). The goal for these kinds of link would be to raise awareness in the advanced language learner of the complexity of the target language while providing stimulating experience in deciphering the interrelationships of complex ideas.

1.3.4 Learner Differences and Strategies

Proponents of hypertext also argue that non-linear information processing allows for different learning strategies within students. Brook, Simutis, and O'Neil (1985 in Stanton and Stammers 1990) have found that there are at least four factors that contribute to learner differences: the learner's intrinsic ability, cognitive style, prior (or background) knowledge, and motivation.

Stanton and Stammers (1990) note that there is a tendency among educationalists to over-categorise learners within the framework of their learning styles (i.e., they try to classify every learner according to a strict set of criteria). We read,

> Caution should be used in approaching the issue of categorisation of learning style. We should not be too rigid in our thinking about styles, and use the available media to allow for the widest individual variations rather than prematurely and permanently classifying learners.

Many other researchers argue persuasively that learner differences such as aptitude, attitude, strategies, and styles play a vital role in the acquisition process (Dickinson 1987). Blair (1982: 229) states that,

> Claims to the contrary not withstanding, in no approach do all learners learn equally well. Whether because of differences in learning strategies, or other factors, in any class some learners perform less well than others.

Other researchers are slightly more specific about why learners are different. Danesi and Mollica (1988) argue that there are learners who learn in synthetic, contextual terms (right brain dominant), and there are those who learn by analysing and formalising functions (left brain dominant). In other words, some students proceed by learning and applying rules, while others listen and practice without analysing why they say what they say.

Stanton and Stammers (1990) argue that non-linear conditions create the type of learning environment that will accommodate these factors within the learner by:

- Allowing for different levels of prior knowledge.
  Students with less knowledge can explore more; students with more knowledge can explore less or as they prefer;
- Encouraging exploration.
  exploration is easy—rather than looking up something in a different book (or perhaps going to a library), users can jump from topic to topic within the same program;
- Enabling learners to see a sub-task as part of the whole task.
  students can move up in a hierarchical task or get information revealing the purpose of the immediate task;
- Allowing subjects to adapt material to their own learning style.
  students who like to browse can browse; students who like to read linearly can read linearly.

Products like *Microsoft Encarta* (an encyclopaedia on CD ROM) exemplify what Stanton and Stammers have in mind here, i.e., one computer-based tool contains the information of ten or twenty printed volumes and includes sound, animation, colour pictures, and video clips—and they accommodate many different information-gathering styles. Apart from the difficulty of reading on a computer screen, these kinds of tool are unquestionably superior to traditional materials for precisely the reasons that are given above.

The opinion that students differ in their learning processes is not in question here. As mentioned in section 1.2.7, it is part of a modern understanding of language learning. The germane question is whether or not non-linear learning provides for different learning styles differently (or better) than do traditional teaching materials.

Non-linear information processing, if designed well, should accommodate a wide variety of abilities, knowledge, and exploratory styles. However, the hypermedia literature is only beginning to describe what students do and to investigate why they do it.

1.4 The Need for Evaluative Research

Quoting Schramm (1977; viii), Flagg traces the development of any new teaching method (in this case, educational radio) as follows:

> Typically a field of this kind which begins without an intellectual father-model begins by trial and error, with many different kinds of contributors, each doing his own thing, before it becomes apparent that all this activity really fits into one field and the different kinds of people in it have something to contribute to one another. Out of this develops a broader and more powerful view of the activity in which they are all engaged. They have been nibbling at related problems; now their activities are ready to coalesce into a field.

Many authors have noted that evaluation of the educational effects of computers is scarce (e.g., Flagg 1990, Byrum 1992). The 'nibbling at related problems' has yet to achieve the kind of coalescence in computers that Shramm foresaw in the field of radio. In short, the CAI field (and CALL in particular) needs evaluation—both for the purpose of improving materials and understanding the effects of those materials on learning.

1.4.1 Reasons for the Gap

Krendl and Broihier (1992 :216) write that:

> ...research on computers and learning is weak in many critical areas—it lacks scientific rigour, conceptual development, appropriate and adequate research methods. Rather than learning

> from and building on research addressing the effects of previous new technologies, such as television, research on computers and learning imitates and replicates the limitations and naivety of such work.

One can argue that there are research gaps in any field—it is the nature of academia and the fact that there are unknowns does not weaken the validity of a field. Krendl and Broihier, however are arguing that there are too many unknowns and that existing research is poorly conceived. Krendl and Broihier are, admittedly, pessimistic in this assessment. Nevertheless, there are a variety of factors that make computer-based materials difficult to evaluate. Flagg (1990) lists six factors which constrain evaluation (and research). These are: *Time, Money, Human Nature, Expectations, Measurement Difficulties,* and *Knowledge.*

Each of these can be explored in the greater context of CAI and also with more specific regard to CALL:

1.4.1.1 Time

Software companies are usually pressed by deadlines and thus produce untested materials. From the practical perspective in CALL, there are few researchers who are not also teachers. Research/evaluation time is usually outside of normal teaching time and is thus limited. Krendl and Brohier advocated longitudinal studies. In their (1992) study, the data collection took place over three years. Even the methodology required by the kind of evaluative research in this thesis requires more time than is available to most language teachers.

1.4.1.2 Money

Research time costs money. Flagg (1990) notes that there is no evidence that formative evaluation saves money for software companies in the long-run—mostly because software companies know that the popularity of materials (i.e., its profitability) does not necessarily depend on the materials' effectiveness. For this reason, software companies do not always undertake formative evaluation. We do know, however, that formative evaluation improves materials in terms of their effects on quantifiable aspects of learning (Krendl and Broihier 1992). What software companies apparently do not realise is that where evaluation takes place, budget costs for pre-production evaluation is low: typically only 2%-3% of total production costs (Flagg 1990), and thus should not be a constraint for software houses.

CALL practitioners must generally rely on their places of work to provide the equipment necessary for both the instruction and the research. Herein lies one of the dilemmas of CALL: the cost of the equipment has not been justified by the research and the research cannot be done without investing in the equipment.

1.4.1.3 Human Nature

Criticism of locally developed materials is not always taken well. Developers often feel that evaluation will stifle the designer's creativity or that over-evaluated materials will be boring. Similarly, programming requires a great deal of time and problem-solving skill. This creativity is not always evident in the end product, where users and colleagues are interested not so much in creative programming as creative instruction. Programmers (or enthusiastic teachers) might be disappointed to discover that their colleagues not only do not appreciate the pedagogic value of their work but also cannot see the considerable effort gone into developing the materials under evaluation.

1.4.1.4 Expectations

Buyers, users, and materials developers often have unrealistic expectations as to what the materials can do. They are disappointed when good evaluation does not produce good materials. Evaluation/research is not a substitute for creativity. The opposite is also true.

1.4.1.5 Measurement Difficulties

Courseware developers often have objectives that are difficult to measure (e.g., creativity) or that require longitudinal studies. Enthusiasts, however, *can* develop hypotheses that are researchable. There are still a plethora of such questions (see below section 1.4.3).

1.4.1.6 Knowledge

Evaluation often is not undertaken because people are unaware of the different methods/philosophies of investigation. CALL research requires high levels of expertise in software and hardware, and also in the Applied Linguistic area of interest (including appropriate research methods). No doubt this fact has negatively affected the volume of evaluative literature.

1.4.2 Rationale for Bridging the Gap

There are two questions that appear frequently in Computer Assisted Instruction (CAI) literature. One queries the effects of certain implementations of computers on learning, and the other examines methods of evaluating materials for improvement. The two are closely related but require discussion in order to create objectives for any research plan.

Flagg (1990) distinguishes between *evaluation* and *research*. The object of evaluation is to assist practitioners in determination of the validity, desirability, and cost-effectiveness of given technological teaching materials. Grant's (1987:119) CATALYST test (designed for teaching materials in general) is an example of evaluation. CATALYST stand for:

- Communicative
- Aims appropriate
- Teachability
- Add-on availability
- Level
- Your impression
- Student interest
- Tried and tested

By using criteria such as *your impression*, and *tried and tested*, Grant is not interested in the effects of materials in terms of some specific, desirable change in learner's knowledge or capability. Here, the emphasis is clearly on the suitability of materials for a specific teaching context. The CATALYST test is designed to assist a decision making process only; it is *summative* (evaluating what already exists in the context of its future 'usefulness') and lies at one extreme of the investigative process.

Pure research lies at the other end of the spectrum. Here, investigators are concerned only with the acquisition of new knowledge, i.e., knowledge for the sake of knowledge (Flagg 1990, citing Unwin & McAleese 1978). Relevance is often determined by current interest in the topic area and usefulness is defined by the extent to which the information enlightens further research.

Formative Evaluation lies somewhere between pure research and summative evaluation. It is evaluation that uses the methodology of research and the objectives of rough-and-ready evaluation tests like CATALYST. Formative Evaluation is a growing and increasingly well-defined subject area in educational technology. Flagg (1990) writes that formative evaluation,

> helps the designer of a product, during the early development stages, to increase the likelihood that the final product will achieve its stated goals. *Evaluation* in this definition means the systematic collection of information for the purpose of informing decisions to design and improve the product. The term *formative* indicates that information is collected during the formation of the product so revisions might be made cost effective.

To a degree, the label on the investigative activity depends on the motivating factors involved and the people doing the activity (i.e., if it is a software developer or television production company, the activity is *evaluation*; if it is an academic institution or research facility, then the activity is *research*). In emerging fields (such as hypermedia), researchers (or evaluators) are interested in similar issues such as effectiveness, appeal, excitement, or program reception. The evaluator, however puts his/her emphasis on the immediate development project. The results of the evaluation determine the success or failure of that project only. The researcher puts his/her emphasis on the effects of the project on a specific educational process. Research might determine the success or failure of an immediate project but it regards objectivity, reliability, and validity as the predominant criteria in the investigative process.

The purpose of formative evaluation is to improve the effectiveness of materials. While this thesis is concerned ultimately with more than just making one set of materials better, it admits to a bias in this direction. The next chapter will discuss the literature of vocabulary and vocabulary acquisition in terms hypermedia. In Chapter 3, we will argue that the materials, while useful in an experimental context, might also be usable in a pedagogical one. This research is not only concerned with measuring the effects of the present materials in certain specific learning areas, but also with making the materials more professional, easier to use, and enjoyable for the student. It is hoped that evaluation of these materials will provide insight for future developers, many of whom are practising teachers. The question "does it work?" is of primary concern but we are also interested in the question "how does it work with which students?".

1.4.3 Descriptions of the Gap

We have said that CALL and hypermedia are only beginning to be discussed in the Applied Linguistics literature. We have, in passing, covered several issues where the literature is weak. There are three key areas that are in need of immediate research. These are: (1) user freedom and control, (2) computer-based feedback, and (3) interface design. Additionally, no discussion of materials is complete without some assessment of affective factors involved with computer use. In particular, anxiety and enjoyment are of interest in a study where many novice users are involved. The following section will examine each of these in turn and then propose a model describing their relationship to each other and how they fit within the general notion of interactivity.

1.4.3.1 Affective Factors

1.4.3.1.1 Anxiety

In most people, the use of computers results in some anxiety. McInerney, McInerney, and Sinclair (1994: 28) have said that computer anxiety is,

> ...an affective response of apprehension or fear of computer technology accompanied by feelings of nervousness, intimidation, and hostility.

The authors have also noted that these responses may include feelings of embarrassment, fear of looking foolish, or fear of damaging computer equipment. These feelings are not foreign to most computer users.

The authors argue that computer experience is one of the primary correlates of anxiety—as experience grows, anxiety diminishes. They add, however, that it is far from certain that users always become less anxious as they gain experience. Indeed, McInerney *et al.* found that the level of anxiety in their students was influenced by other factors including: (1) what the users were being asked to do (was the computer interaction structured or unstructured?); (2) the purpose of the computer use (was there a threat of evaluation?); and (3) the perceived goal of the computer interaction (was it performance evaluated against other students or was the focus on the intrinsic value of learning?).

McInerney *et al.*'s study notwithstanding, the wide variety of machines, operating systems, interfaces, and program content make generalisations about the causes of anxiety extremely difficult. For example, when a user changes between operating systems (or when they are subjected to technological changes that arise through software or hardware 'upgrades'), even experienced users may suffer anxiety similar to that of novices.

The variability within and among users themselves is another restraint on generalisations about the causes of anxiety. Students react differently to computers not only because of the computers (and the differences between hardware, systems, and programs) but also because as individuals some are naturally drawn to computers and technology, some avoid contact with technology altogether, and some are naturally prone to be anxious about almost anything (see Crable, Brodzinski, Scherer, and Jones, 1994 for a discussion of individual differences in computer anxiety and Ely 1988 for a discussion of student personalities and classroom activities). User age and gender have also been investigated as a correlate of computer anxiety. Massoud's research (1991), for example, suggests that there is some relationship between gender and attitudes towards computers (males had a more positive attitude towards computers in that study and there was no link between age and attitude). The number of potential variables is large. When assessing the effects of anxiety on specific materials, it must be kept in mind that anxiety is dynamic and potentially recurring; it depends not only on users' experience but also on the personality of the student and the task she or he is asked to undertake.

Reading while listening (and while clicking with the mouse) is not something in which most students have extensive experience. The students will potentially experience anxiety if two conditions are evident: (1) students are novice computer users, and (2) students have had little practice reading while listening.

1.4.3.1.2 Enjoyment

Enjoyment of computer materials is another affective factor that will be examined in the design described in Chapter 3. Enjoyment (or enjoyableness as Green (1993) calls it), has not seen as much attention as anxiety. Green (p. 1) has said that,

> ...it is surprising that almost nobody seems to have actually asked students to rate the extent to which they enjoy different classroom activities.

Green asked students to rate both their enjoyment and the effectiveness of classroom activities. Effectiveness was operationally defined as: *the opinion that the activity would help students to become more competent in English.* He found a correlation between reported enjoyableness and perceived effectiveness of various EFL techniques.

The relationship between enjoyment and effectiveness is not clearly understood. It is not known, for example, whether certain methods are effective because they are enjoyable or enjoyable because they are effective. Green (1993: 8) argues that the relationship between enjoyableness and effectiveness is "circular and mutually reinforcing...that enjoyableness enhances effectiveness, and that the belief that something is effective tends to make it more enjoyable than it would be otherwise." No doubt, in Green's research where both effectiveness and enjoyableness were assessed by the subjects, this circular relationship would be reinforced.

Green was referring primarily to traditional (i.e., non computer-based) methodologies in his study of enjoyment and effectiveness. The present study, while primarily concerned with the effects of computer use on vocabulary acquisition, is also interested in enjoyment and other affective factors. As is the case with many subjective factors, 'enjoyment' could be operationally defined in many ways—including the results of psychological testing, interviews, or both. On the other hand one could take a more 'rough and ready' approach by first, as Green did, simply asking the students if they enjoyed it, and second, asking them how useful they thought the activity was for learning English. Where the primary issue of the research is a study of enjoyment, this approach may be unsatisfactory. However, in the context of evaluating materials designed to promote vocabulary acquisition, and where the issue is the suitability of these materials for a self-access centre, it is enough to know how students reported their enjoyment and whether or not they deemed the procedure effective for learning English. The results from such an investigation should be considered as a preliminary step—one that might lead to a more comprehensive investigation in the future.

1.4.3.2 User Freedom

Freedom for the user is often seen as one clear advantage CALL holds over traditional teaching materials (cf. section 1.3). However, Sciarone and Meijer (1993) observed that students will take advantage of the system if given too much freedom.

The authors imply that there is a difference between 'freedom to cheat' and 'freedom to use alternate tasks to accomplish a worthwhile goal'. The latter is considered a desirable feature while the former was found by these researchers to produce undesirable effects on learning (defined in their study as performance on cloze tests).

Hypermedia can easily be made to allow students to wander freely, and students often use the materials without close supervision. The issue of user freedom, therefore, is of crucial interest for hypermedia designers who want students to work autonomously but not without purpose.

1.4.3.3 User Control

The notion of user control is closely related to exploratory learning and is a function of user freedom. Much has been said in the field about allowing students to go at their own pace, sequencing their exposure to language according to their own preferences, and motivating students through greater freedom. However, Borsook and Higgenbotham-Wheat (p.13), based on their observations of students using computers note that,

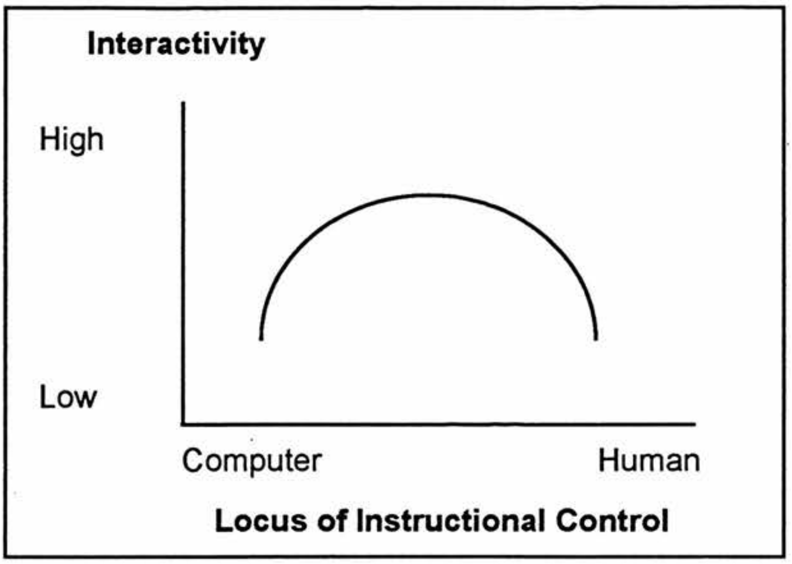
> ...total learner control is beneficial only to those who are already somewhat knowledgeable about a domain or who are generally high achievers....It appears that for most learners, handing them over too much control is like giving them "enough rope to hang themselves with".

As a general principle, this statement is probably useful but it does not say enough about the fact that there are different kinds of materials for different purposes. Certainly, if the materials developer wants students to obtain a specific skill from a computer lesson (e.g., a medical student obtaining the skill of understanding patients' complaints in English) then providing completely unguided materials would be counter-productive (the student might spend so much time investigating ancillary materials that they never get to sections covering the target skill). On the other hand, in this author's experience, students often come to materials with their own agendas and get frustrated when they are forced to do things that are not germane to their own needs (e.g., a students getting information for a project from a CD ROM encyclopaedia does not want to be forced to take an on-line quiz in order to use the resource). Thus, the level of user-control depends largely on the purpose of the materials and the needs of the user.

Borsook and Higgenbotham-Wheat's describe interactivity in terms of the *locus of control*. They argue that hypermedia documents with total user control are like one-sided conversations—they are not interactive. They propose an explanation of interactivity with the following graph (p. 14):

Figure 1.4-1

Interactivity v. Instructional Control



In this model, interaction occurs when there is a balance of control between the user and the computer. Again, the model is useful if used in the context of specific types of materials and assuming that the user has some computer skill. The skill of the student (which is usually a function of their previous computer experience), will affect the level of interaction in any given set of materials. This, of course, greatly depends on the type of task involved. Tasks requiring typing skills, for example, will be less interactive if the student is unfamiliar with the keyboard because the student (apart from feeling frustrated) will respond slowly to the computer's prompts (see van der Linden 1993 for a discussion of keyboard difficulties).

High interaction is generally considered a desirable element in language teaching. The computer can provide a degree of interaction, but can never be expected to replace the kind of interaction a good teacher can provide. For example, a teacher has an implicit understanding of a student's individual learning style and tailors the instruction for the types of learner in his/her class. Good teachers can 'sense' the correctional needs of the student—knowing when correction would be beneficial and when it would produce anxiety—and respond accordingly. Good teachers react to students on a personal level with a degree of warmth, compassion, empathy and sympathy simply not available with human-machine interaction.

Experts in artificial intelligence (AI), in the hopes of providing better computer-human interaction, attempt to create programs that will 'model' different types of student. The problems with student modelling are many because of the complexity of students and the differences between individuals. To exemplify the difficulty, we might consider Boyd and Mitchell's (1992: 24) argument that even the identity of the student is problematic:

> ...a given "mechanical" individual "John Brown" may actually operate in a learning conversation as one of several personae or participant...individuals (e.g. enthusiast, constructor, critic, integrator)...a learner (or teacher) may switch personae during the course of instruction so that the assumed "personal style" may not be the one in use at all at some given phase of instruction.

Taking only this one example, the enormous complexity required to mimic 'real' interaction is self-evident. When one considers all the other ways in which a teacher 'understands' his/her student (knowledge of a student's level, culture, world knowledge, etc.) it becomes difficult to imagine a computer program that would interact with a student *even* as well as a human teacher does—never mind replacing that teacher.

1.4.3.4 Feedback

One area in which the computer is seen as much weaker than traditional classroom-and-teacher scenarios is that of direct feedback. Many hypermedia programs make use of drill and practice, cloze, flashcards (or other activities that fit into the 'guess and check' category of learning) as ancillary activities that branch from the main program. In computer-based drill and practice activities, good feedback is vital but difficult to use effectively. More importantly, the definition of 'good' or 'effective' computer-based feedback is only beginning to be explored.

Van der Linden (1993) is an example of such preliminary research. The author carried out research exploring students' navigational patterns (using a log system) in an attempt to determine the effects of different types of feedback. Van der Linden argues that computer-based feedback is fundamentally different from (and arguably superior to) teacher-based feedback in that it can (1) be provided for every incorrect response, (2) be given immediately, and (3) requires self-correction by the learner. Findings from the van der Linden study suggest that the type of feedback is important and the way in which it is presented may determine the likelihood of it being used (for example, she argues that the length of the text on screen must not be more than three lines). However, individual differences in the way in which the feedback was used also determines the usefulness of the feedback. Some students used the facility extensively; others ignored it completely. Again, the issue comes back to user freedom and careful observation of the way in which students use that freedom.

Van der Linden's research suggests that logging systems can be used to examine the strategies of users with regards to computer-based feedback. In this way, it is similar to the Sciarone and Meijer study. The two issues are closely linked. Students are generally free to use various options (including feedback) but *why* and *how* students utilise (or fail to utilise) these facilities should be the focus of more research in the near future.

Accurate and informative logging of students' actions while using hypermedia materials is possible (and usually quite easy to provide) yet many research projects make no use of a logging facility. If used together with questionnaires, interviews, or 'think-aloud' protocols, computer-based logs could be used to learn much about student behaviour (both as individuals and as groups), cognitive processes, and student preferences.

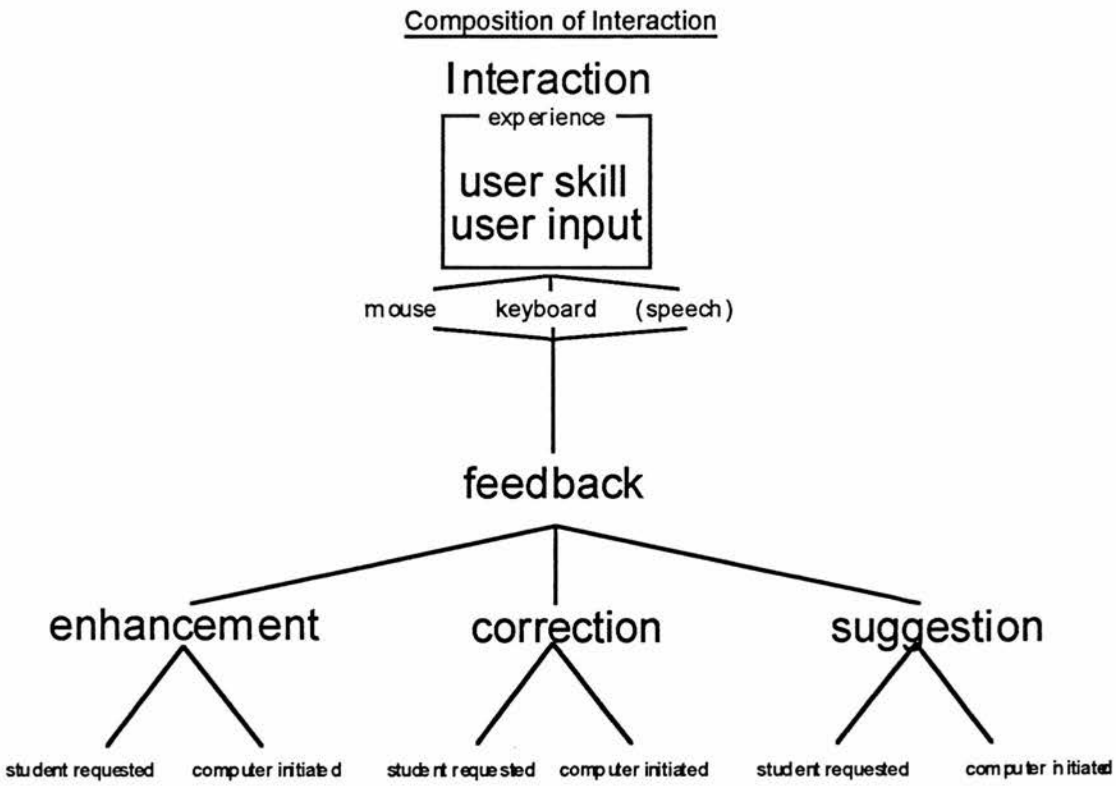
1.4.3.5 Interface Design

Interface design of hypermedia materials (and here we are talking not about operating system design but the way specific materials require users to interact with it) is another area that is still at the beginning stages of description and evaluation. Some useful discussions of non hypermedia interfaces and design do exist, however. Norman and Draper (1986) approached the issue of interface attempting to provide a philosophical support for specific design implementations. Their volume contains several articles aimed at the psychology of the interface (e.g., Laurel, Bannon) as well as others that take a more 'how to' perspective (e.g., Mark). Other books focus almost exclusively on the latter. Shneiderman (1987), for example, provides principles for creating an attractive and comprehensible interface (including ideas for the use of colour and menus) while offering useful descriptions of terminology and hardware. Description of terms and discussion of available hardware and software can also be found in Nickerson (1986). The first problem with these handbooks is that they are not aimed directly at foreign language teaching. Non-English speaking users have the additional potential difficulty of not being able to understand menu options and instructions because they are written in English. The other problem is that in the years since these books were written, the possibilities of the interface have changed rapidly—especially with the advent of *Windows* and hypermedia. The works just mentioned should not be discarded altogether (indeed, in the case of Shneiderman in particular, many of his suggestions are still invaluable), but they are dated. One would expect that similar works will soon appear for graphical interfaces but until then, materials must be evaluated on a case-by-case basis—in the hopes that findings might contribute to general principles of hypermedia design.

Many studies have concentrated on interface evaluation through the observation of materials in use (e.g., Byrum 1992). This is a kind of formative evaluation (as discussed in section 1.4.2) and it improves materials. It is necessary in a pedagogical context. However, conclusions from this type of study, while helpful are usually non-theoretical. Studies with university students have found, for example, that an overly simplified interface will encourage the perception that the materials themselves are simple and students may feel patronised as a result (Arnold *et al.* 1994 and also Spiro and Jehng 1990). The field would benefit from more comprehensive and data-driven studies investigating what students expect, what disappoints them, what they feel assists them, and how much they feel they benefit from specific types of hypermedia-based materials. A consolidation of such data with existing and emerging theories of language learning is essential for the continued development of more effective and understood materials.

Feedback, freedom, and interface design work dynamically together to determine the degree and kind of interaction in a program. The following diagram is an illustration:

Figure 1.4-2



A few points need clarification. First, as has been said, users' computer skill is dependent on experience (and the related issue of anxiety). Different kinds of physical interaction require more or less experience. Feedback is subsequently dependent on skill and students' input into the computer. This diagram shows three types of feedback: enhancement, correction, and suggestion. *Enhancement* describes multimedia or hypertextual additions to a text (e.g., reading while listening, narrated video or animation, on line tools, etc.). *Correction* is the computer's response to a user's mistake and *suggestion* is a hint the computer might give before correction in order to elicit the correct answer. All three characteristics can be computer or user controlled. As has been said, it is the degree to which one or the other has control that determines the interactivity of a program.

1.4.3.6 Effects on Learning

User freedom, feedback, control, and interface design work together to determine the level of interactivity. How these factors interact with each other is still the subject of debate. Indeed there is much preliminary research to be done in this area. In addition to dynamic relationship between user and materials, very little is known about the effects of each of these factors on specific areas of learning in general and language learning and acquisition specifically. Several questions come to mind: First, in general:

- What makes a physical interface difficult?
- What makes a screen interface difficult?
- What graphical images and metaphors are best understood by the user?
- How necessary is experience with materials (even those that are propounded to be 'easy to use')?
- What do users expect from modern materials?

Second, in language learning:

- What effect, if any, do hypermedia features like sound, video, animation, etc. have on language learners' vocabulary acquisition, comprehension, syntax acquisition, and pronunciation? What can hypertext do to improve students' cultural knowledge of the target language and world knowledge in general?
- Do language students enjoy the addition of hypermedia features?
- Do language learners think they learn when hypermedia features are added and if so, what do they learn and from which features?

1.5 Summary

Since the advent of computer assisted instruction in the 1950s and 1960s, there has been a desire to see computers as more than just a tool that accesses information quickly. Early enthusiasts usually envisioned the 'teaching machine' as the ideal for which to strive. While few would deny the practical uses of computers as word processors and data managers, it is slightly more controversial to propose that a computer can be an autonomous teacher. Unfortunately, there is still only a limited body of research reporting the educational benefits of specific uses of computers in language learning.

What can we conclude about the prospects for CALL? In the 1980s, hardware costs and availability permitted educationalists to do preliminary investigations into the cognitive processes involved in human-computer interaction. The possibilities of the medium opened up an inspiring and hitherto unexamined area of research. In 1984, Thomas (1984: xvi) wrote that,

> We stand on the threshold of a new age brought about by new technology. The extent to which that new age represents progress in human productivity and enjoyment or merely a change in

> fashion without real progress depends heavily upon the work of investigators...

The investigation envisaged here is still in progress. Moreover, unlike most other areas of Applied Linguistics, the investigative process is complicated in CALL by the ever-changing nature of the medium. Research that is interesting over a period of two or three years can quickly lose its relevance. If a researcher puts too much emphasis on questioning the effects of a specific technology, he or she is likely to find that by the time the research is complete, the original question is moot. Researchers must therefore be careful to ask questions whose answers will apply to the next generation of hardware and software as well as the immediate one.

Computers offer researchers unique opportunities. Not only can research test the effects of new types of learning (such as non-linear and hypermedia) but there is more control over research variables (like the variability within the teacher) than is possible with traditional classroom-based methodologies. Still, the methods for evaluating computer-based methods must be carefully considered and delineated. Appropriate and relevant linguistic-based hypotheses must be posited and tested. The next chapter will discuss vocabulary and vocabulary acquisition as an area that might be assisted by hypermedia materials.

Table of Contents: Chapter 2

Chapter Two
Second Language Vocabulary: History, Acquisition, Instruction and the Possibilities within CALL

Introduction

This chapter will discuss the literature of second language vocabulary acquisition with regards to its historical significance and current status. Section 1 will briefly define some key terms in the field, paying particular attention to acquisition and learning, the meaning of 'word', what it means to know a word, and the active/passive knowledge distinction. Section 2 will examine the progress of lexical research and teaching over the past fifty years including a discussion of its importance in the theoretical context over time, changes in scholarly interest, and an overview of the mental lexicon. Section 3 will begin with a slightly more narrowed discussion of incidental vocabulary acquisition. Included in this section will be an examination of first language and second language research in vocabulary acquisition through reading, followed by first and second language research in acquisition through reading and listening. The last section (section 4) will briefly discuss the possibilities of using the computer as means of enhancing vocabulary acquisition through the addition of spoken text reading materials.

2.1 Clarifications

2.1.1 Acquisition and Learning

The distinctions made several years ago between acquisition and learning were an attempt to further clarify the familiar distinction between subconscious and conscious learning (Wode 1981). Krashen (1977) argues that what is learned enables learners to 'monitor' what is acquired. 'Acquired' language produces speech while 'learned' language corrects, edits, and otherwise 'monitors' the learner's speech. The model gives little attention to the place of vocabulary (learned or otherwise) in the acquisition and learning process, despite well-documented and extensive debates about Krashen's monitor hypothesis. Researchers interested in vocabulary have seemingly returned the favour. Meara (1992), Arnaud (1992), Carter (1987), McCarthy (1984) and many others, either give no mention to the monitor theory in relation to vocabulary or they simply use 'acquisition of vocabulary' and 'learning of vocabulary' without differentiation. Wode (1981:59) writes that,

> ...to me, the main difficulty [with the monitor model] is that I fail to see why learning a language and why correcting one's speech production, i.e. editing it, should be as directly linked as implied by Krashen's conception of the monitor.

In the case of vocabulary, there is even less in the monitor model to suggest how what is 'learned' (in the monitor model sense) affects what is 'acquired', and whether or not the learned knowledge interacts with the acquired knowledge in the way that the monitor model intimates about other aspects of language such as morpheme and syntax acquisition.

This is not to say that vocabulary research makes no distinction between conscious and subconscious learning. On the contrary, concern with comparisons between incidental (or subconscious) vocabulary acquisition (e.g., Hulstijn 1992) and learning vocabulary using specific (conscious) techniques (e.g., Carter and McCarthy 1988; Carter 1987) demonstrates that the conscious/subconscious distinction is not only recognised but also of primary concern in the field of second language vocabulary studies. For this reason, 'acquisition of vocabulary' and 'learning of vocabulary' will be used interchangeably throughout this thesis (and compared with the following section). However, the distinction between incidental acquisition (from reading, for example) and active learning (through active techniques like memorisation and drills) will be maintained. Arnaud and Béjoint's (1992: xi) simple description of the study of *vocabulary learning* should be kept in mind:

> *Vocabulary learning* is a mental activity which, like all mental phenomena, can be viewed from many angles. [The question is:] How do foreign words and other lexical units find their way in the learner's mind and how are they organised there?

This question forms the basis of most research in vocabulary acquisition. The 'angle' of view of this thesis will be more fully described in Chapter 3.

2.1.2 Words

For the purpose of this study, a 'word' can be defined as a lexeme, i.e., an abstract unit that can be realised through different word forms and consists of one dictionary entry (Carter 1987). Lexemes can be one or more words, the parts of which do not separately convey the meaning of the whole (*Concise Oxford*

Dictionary 1990). The terms *word* and *vocabulary item* and *lexical item* will be used interchangeably. Similarly, the terms *lexis* and *vocabulary* will be used interchangeably except where a distinction is noted.

2.1.3 'Knowing' a Word

When discussing vocabulary acquisition we must first determine what we mean by 'acquiring' or 'knowing' a word. Intuitively, one might be tempted to say that *knowing* a word is a binary choice—you can define the word or you cannot. The manner in which standard dictionaries list and define words might encourage this view. The definitions in even the best dictionary do not, however, contain more than a fraction of the possible information about the word (Aitchison 1987). Krashen, (1989: 446) writes that,

> words are not learned all at once when they are seen in context; rather, word knowledge grows in "small increments." At any given time, there are words we know well, words we do not know, and words in-between. To increase our vocabulary we need to learn to...tolerate some vagueness, vagueness that is reduced bit by bit as we read more and encounter unfamiliar words more.

Similarly, Twaddell (1973: 72-3) notes that,

> ...within our native-language vocabulary at any given time there are words that are familiar or only partly familiar. And we also "know" a very large number of words with various degrees of vagueness—words which are in a twilight zone between the darkness of entire unfamiliarity and the brightness of accurate complete familiarity.

Understanding goes beyond knowledge of a word's meaning. It extends to non-semantic information as well. Richards (1976), in his study of the effects of theoretical hypotheses on language teaching, notes that there are at least seven aspects to 'knowing' a word (see also Oxford and Scarcella 1994 for a similar definition):

1. knowing a word means knowing the degree of probability of encountering that word in speech or print. For many words, we also "know" the sort of words most likely to be found associated with the word;
2. knowing a word implies knowing the limitations imposed on the use of the word according to variations of function and situation;
3. knowing a word means knowing the syntactic behaviour associated with the word;
4. knowing a word entails knowledge of the underlying form of a word and the derivations that can be made from it;
5. knowing a word entails knowledge of the network of associations between that word and other words in the language;
6. knowing a word means knowing the semantic value of the word;
7. knowing a word means knowing many of the different meanings associated with the word.

This study will focus primarily on number six. That is, the experiment described in the next chapter will be limited to measurements of a single semantic value (the meaning of the word as it appears in the experimental context) of a target word.

2.1.4 Active and Passive Word Knowledge

For the purposes of this experiment, we will accept the traditional, dichotomous distinction between *active* and *passive* word knowledge as unproblematic. By way of definition, words known passively are words that are recognised when they are encountered but where there is no capacity to recall them spontaneously or in association with other words. Conversely, words known actively can be recalled at will or recalled in association with other words. If word knowledge lies on a continuum from less known to fully known, the continuum exists in actively known words. There is no need, in this thesis, to reject the notion that passive knowledge is something fundamentally different from active knowledge. Meara's (1990) delineation of active and passive lexical knowledge in terms of graph theory (active words have two-way associations to other words and passive words have a one-way association) is a useful and sufficient explanation for present purposes.

2.1.5 Learning Strategies

The way in which learners read and the strategies with which they learn is characterised by variability both within learners and between learners. By way of definition, Green and Oxford (1995: 262) have said that,

> language strategies are specific actions or techniques that students use, often intentionally, to improve their progress in developing L2 skills.

Research suggests that learners are aware of many of the strategies that they use and that they will adapt their strategies according to the requirements of the task. Less succesful language learners are less able to choose appropriate strategies and to link them in useful patterns (Green and Oxford 1995; O'Malley and Chamot 1990).

Language learning strategies take different forms for different students for different reasons. Students vary according to gender, age, affective factors like motivation and anxiety, and ethnicity; their learning strategies parallel these differences. O'Malley, Chamot, Stewner-Manzaneres, Kupper, and Russo (1985) delineated various learning strategies in terms of students' language acquisition activities at various levels of language proficiency. Citing Wittrock (1975), O'Malley *et al.* note that learning is most effective when it involves processes that link new meanings with old meanings. Furthermore, comprehension is highest in a reading environment when motivation and willingness to invest effort is high, and when the reader has an ability to attribute success or failure to the effort (O'Malley *et al.* 1985). Effective learning strategies are characterised by student involvement in or dedication to the task. Direction or systematicity also seems to be evidence of good learning strategies.

In terms of computer-based reading, it is systematicity (or lack of systematicity) in the patterning that is most apparent by the data produced by computer-based log files. This kind of data shows that users work their way through computer-based materials in various ways including linearly (from beginning of the lesson/document to the end), repetitively (working through certain segments, then doing it again—the size of the segment depending on the student), starting and stopping (the user gets to a certain point then 'gives up'), or randomly (no *obvious* pattern is is evident to the researcher). Using patterns alone, it is difficult to determine why students do what they do (cf. Cohen 1987). Nevertheless, it is possible that these patterns reflect the strategies of effective/ineffective listeners. O'Malley and Chamot (1990), for example, report that ineffective listeners characteristically are not distracted by encounters with unkown words, they simply infer from the context (as much as possible) and then continue to listen without breaking their concentration. However, ineffective listeners in the same circumstances, tend to stop listening and lose their concentration without being aware of their inattention. O'Malley and Chamot argue that effective learners recognize when their attention is lapsing and are more likely to redirect themselves to the task of comprehension. Similarly, generally speaking, effective listeners tend to use both a 'top-down' and 'bottom-up' approach while poor listeners use only a bottom-up approach (i.e., poor listeners approach comprehension with a word-by-word strategy while better listeners are more likely to take 'chunks' of oral text and look for the meaning of the chunk). It is certainly possible that patterns in log files exhibiting start and stop patterns, random patterns, or word-by-word patterns, are evidence of the strategies of weak listeners who are easily distracted while systematic patterns, including repetition of larger segments, linearity, or a combination of systematic patterns are evidence of effective learners who purposefully (cf. O'Malley and Chamot 1990) utilise top-down and bottom-up strategies.

It is important to stress that the data collected from computer-based logging systems describe actions only, not strategies. Ideally, given enough time and cooperative students, a researcher might combine these data with various forms of think aloud observations (see Cohen 1987 for a discussion of various protocols) and with test data to measure performance. In the present study, the design permits only comparisons of navigational patterns with vocabulary test results.

2.2 Vocabulary in this Century

2.2.1 History

Vocabulary goes in and out of fashion depending on prevailing linguistic (or applied linguistic) theories and corresponding trends within the language teaching profession. Richards (1976: 77) writes that,

> The teaching and learning of vocabulary has never aroused the same degree of interest within language teaching as have such issues as grammatical competence, contrastive analysis, reading, or writing, which have received considerable attention from scholars and teachers.

Almost ten years later, McCarthy (1984) similarly wrote that the treatment of vocabulary in many coursebooks at the beginner and lower intermediate level is regrettably insufficient. Apparently, the state of vocabulary instruction changed little in those ten years.

The lack of recognition and interest in vocabulary was established by Meara (1982), who discussed the neglect while listing key areas that might benefit from immediate research. However, going back in time to the 1920s and continuing to the late 1960s, vocabulary, while not the centre of focus in language learning, featured in theoretical and pedagogical discussions. For example, Halliday, McIntosh and Strevens (1964) call the interest in vocabulary a 'high water mark' in linguistic sophistication. He notes that teachers welcomed data from word count investigations because they provided a rationale for the choice of classroom materials. Despite the fact that word frequency lists were almost always based on written texts, teachers intuitively recognised the validity of claims that students need a 'core' vocabulary. The particular words included in this core could be justified using empirical methods like word counts in a particular textual corpus. The work of Halliday *et al.* (and many others of the time) emphasised the importance of vocabulary instruction, placing it prominently beside other (what were in their view) requisite aspects of second language learning like grammar and pronunciation.

Howatt (1984) traces the interest in vocabulary in this century to Michael West, who in the 1920s determined to teach reading as a first priority "even if this meant the relative neglect of the spoken language." (Howatt 1984: 246). Reading texts could be improved by controlling the kind of vocabulary to which young readers were exposed and the manner in which it was presented. Control and management of new words was achieved through selection of 'common words' (ibid., p.247); through distributing new words more sparsely throughout a text; and through indicating target words to teachers and students by including lists in the reading text. Howatt notes that West incorporated lexical control into a strategic, staged approach to second language reading. West's first stage in reading instruction was a careful introduction of new words, the second was to hold vocabulary constant while concentrating on the development of skills, and the third stage was to incorporate skimming and scanning techniques into second language reading. Howatt goes on to argue that West's strategic approach did not fit into the intellectual atmosphere of the time, where, as mentioned above under the writings of Halliday *et al.*, teachers were devoted to the observance of theoretical linguistic principles of careful (Howatt calls it 'obsessive') attention to the detail of vocabulary selection, structural organisation and grading. West's work did, however, fit in well with theoretical interests in vocabulary. His *General Service List of English Words* (1953) reflected an undeniable interest in vocabulary among some linguistic theorists and teachers of the time. Despite the interest in word frequencies, however, vocabulary was still very much at the background of language learning theory. Meara (1982:

103) notes that "the objective of this type of research [i.e., vocabulary counts] is to decide what words are to be taught, not to find out how words are actually learned.". It is possibly for this reason that acquisition theorists are not overly concerned with frequency lists—although teachers recognise the usefulness of word lists based on additional factors such as familiarity, need, utility, etc. (Richards 1974 and Honeyfield 1977).

In the classroom, the first interest was still structure. For example, Carter and McCarthy (1988) note that works like Fries' (1945) *Teaching and Learning English as a Foreign Language* make almost no mention of lexis as a part of second language pedagogy. Similarly, in Rivers' (1968) *Teaching Foreign Language Skills* the dominance of structure and the treatment of vocabulary as a 'by-product' of structural instruction can clearly be seen. The by-product approach was only supplemented when the learner was in need of vocabulary for production, expansion of recognition skills, or for learning in specialised areas of expertise (Carter and McCarthy 1988).

Carter and McCarthy cite several perceived reasons for the playing down of vocabulary (based on Rivers (1968)).

1. it is difficult to predict what vocabulary learners will need;
2. over-concern with vocabulary gives learners the impression that language learning is just the accumulation of words;
3. hesitancy of recall can result from excessive vocabulary teaching;
4. first-language acquisition proceeds with a small vocabulary until structural patterns are mastered.

These factors were to provide the rationale for the de-emphasis of vocabulary well into the 1960s. In the following two decades, however, vocabulary was disregarded because of developments in second language acquisition (SLA) research. SLA theorists tended to take an exclusive interest in morpheme and syntactic acquisition while overlooking lexical acquisition and teaching.

Laufer (1986a) examines the lack of interest in vocabulary by following the attention that 'classical' language acquisition books, written in the 1970s and early 1980s pay to the question of vocabulary instruction and acquisition. In general, textbooks and commentaries aimed specifically at language acquisition devote only cursory attention to lexis. Laufer notes that in Hatch (1978), Dulay, Burt, and Krashen (1982) and others, it is as if

> language acquisition meant only the acquisition of morphemes and syntax....The content of the above mentioned studies seems to suggest that vocabulary has not been a good source of inspiration for investigators of language acquisition, error analysis, or interlanguage (Laufer 1986a: 69-70).

Laufer wrote at a time when traditionally perfunctory attitudes towards vocabulary began to change and the contemporary emphasis of lexical importance began to emerge.

2.2.2 Traditional Vocabulary Instruction

Early vocabulary instruction often consisted of lists of pairs. These lists contained either a word with a synonym or a word in the L1 that was supposed to 'equal' the word in the L2. The text was often accompanied by a graphic or some other mnemonic representation. Some studies have shown that in the initial stages of learning a language, large numbers of words could be learned quickly and efficiently by simply memorising pairs (Carter and McCarthy 1988). Others, (e.g., Gairns and

Redman 1986) comment that in the early stages, pair matching and repetition give the students the opportunity to manipulate the oral and written forms of language items and many students get a strong sense of accomplishment from these kinds of activity.

Further on in the development of the second language, however, this approach is unsatisfactory. Twaddell (1973: 66), in a discussion of the appropriate learning stage to seek vocabulary expansion, writes that,

> A gloss is better than nothing, but it is no more than a starting point. The least valuable information about a word is one word in another language. To treat pair-matching as a learning objective or as a testing device violates what we know about polysemy and differences among languages.

2.2.3 Changes in Instruction

By 1984, writers began to accept the importance of vocabulary and context more wholeheartedly. McCarthy (1984 :12), says "...a steadily growing amount of work is beginning to challenge assumptions that have relegated vocabulary teaching to a secondary position in the priorities of language teaching". Despite the growing interest in vocabulary, however, McCarthy points out that the some of the most popular coursebooks of the 1980s (e.g., *Kernel One (1978)* and *Streamline Departures (1978)*) emphasised the grammatical purpose of any new lexical items. Drills and other exercises in these types of book are designed primarily to reinforce new grammatical paradigms. Coursebook authors are reluctant to increase lexical load (in order to increase grammatical learning) and vocabulary is expected to be acquired incidentally (note that there were several variations to this method—

students might be asked to identify new words, define them, then use them in their own sentences, for example). The result, however, was an obsession with words in isolation (McCarthy 1984) and a *de-emphasis* on the way in which words are used. The structure-focused coursebook tends to ignore that fact that ideas are often reiterated using related words. McCarthy offers the following example:

A. Were you angry?
B. Yes, I was absolutely furious.

As opposed to a more structurally-minded approach:

A. This engine is useless.
B. Yes, it is.

The first example is more reflective of the way in which vocabulary is actually used, i.e., vocabulary does not exist in order to fit some prescribed syntagmatic pattern but to perform a function of informational content and possibly for emotive and emphatic purposes.

McCarthy is concerned here with making the learner aware of the 'company words keep'. New words must not be presented simply to make the use of grammatical paradigms possible; they should be presented with relevant collocates that make sentence and text construction more obvious than intuitive (McCarthy 1984).

Studies in collocation also represent a change from the 'word-in-isolation' thinking. Collocation research (e.g. Brown 1974) emphasises that words are more likely to appear with some words than others. Methods that emphasise the collocationary characteristics of words tend to suggest that students concentrate on what comes before or after the word in question. The purpose of emphasising the collocationary characteristics of words is to allow students to get a 'feel' for what sounds 'right' (Brown 1974). This 'feel' results from having encountered words many times with other words that native speakers would naturally collocate.

Twaddell (1973) argues that since native speaker reading skills and comprehension strategies compensate for most lapses in first language lexical understanding, second language learners must be encouraged to use the skills of guessing and inferencing that they would use in their own L1. Fostering, developing, and transferring the guessing skills that have been acquired in a learner's L1 is the learner's (and teacher's) objective (Twaddell 1973). *Context*, in this view (providing that learners can understand the context) provides most of the information necessary for learning new words. Learners' use of context and the methods of encouraging correct use of context are crucial for vocabulary acquisition to take place.

Laufer (1986a) argues that the heightened sense of lexical importance that arose in the early 1980s came about because developments in linguistic studies of semantics are naturally followed by a corresponding interest in lexis in applied linguistics. Furthermore, she writes that the view of language has changed from one that is abstract and idealised to one that focuses on what is functional and social, i.e., "the interest has shifted from sounds and structures to meaning, discourse and speech acts. Since all these involve an adequate use of words, an interest in words is bound to develop." (p. 71). What the learner can do with lexis is as important as what s/he could do with grammar. Carter and McCarthy (1984: 14) say that, "The key to a new approach to vocabulary teaching lies in the examination of the syntagmatic and paradigmatic relations of collocation and set between lexical items." Vocabulary knowledge is now seen as a pivotal feature of language—it is assumed that without an adequate lexical knowledge, the learner can accomplish very little linguistically.

2.2.4 Implications

There were several developments in the early 1980s that demonstrate the general acceptance of principles similar to those set out by Twaddell in section 2.2.2 above:

- mistakes in syntax were more tolerated in pedagogical theory due to the increased understanding of interlanguage as a natural developmental process (e.g., Selinker 1972, Corder 1967, Twaddell 1973);
- reading began to be looked at as the primary means of access to a broader vocabulary (e.g., Elley 1989, Gairns 1986, Krashen 1989);
- the concern grew for advanced learners and their lexical needs (e.g., Laufer 1986a);
- communicative language teaching emphasised fluency over accuracy; and lexis over syntax (e.g., Carter 1987, Laufer 1986a citing Widdowson and Brumfit 1981 and as discussed below).

This is not meant to be a comprehensive list nor does it imply that these principles were not evident in some form in previous years. But as a whole, the list summarises a change in the perception of the importance of vocabulary—both in how it is taught and in how it is learned.

Following the development of communicative theories in the 1980s, researchers and teachers based the definition of 'important' as much on notions of the effects of improper usage of the area in question as on perceptions of what is traditionally done (and therefore 'right') in the teaching of a language. In keeping with this, Carter (1987: 145) says that, "Mistakes in lexical selection may be less generously tolerated outside the classroom than mistakes in syntax." Laufer notes that the general change in the notion of 'importance' comes from a de-emphasis of 'accuracy' and a rising emphasis on 'fluency'. She says (1986a: 72),

> If fluency means the ability to convey a message with relative ease and comprehensibility, then it is vocabulary correctness and adequacy that matter more than grammatical accuracy....It is reasonable, therefore, to assume that those who advocate the supremacy of fluency over accuracy will also realise the supremacy of lexis over grammar.

Syntax, from this perspective, is not as important as vocabulary because unconventional uses of syntax does not result in as much confusion (or possibly offence) as mistakes in lexical choice. From this point of view, what 'works' and what does not work are primary. Taken to the extreme, one might argue that if a learner can communicate more with a large number of words spoken without a recognised syntax than with well-constructed sentences comprised of poorly chosen words, then the former is a more desirable condition. Of course the optimal state is a large number of words in well-constructed sentences, but this is not possible at every stage in the language learning process. Nevertheless, learners themselves constantly tell us that without adequate vocabulary, there cannot be communication or comprehension (Laufer 1986a). Vocabulary instruction should now be seen as an essential element of any language pedagogy. The next section will examine the current understanding of how vocabulary is perceived by researchers and how it is acquired by language learners.

2.2.5 The Importance of Vocabulary

2.2.5.1 The Relationship between Vocabulary and Comprehension

Nation and Coady (1988) argue that vocabulary is the most consistent factor in predicting readability of a given text, i.e., the difficulty of individual words in a text is a more certain determinant of readability than the structure. A word's difficulty is usually defined in terms of its frequency and length. Other measures include a word's associative value (does one word call up other words) and concrete *v.* abstractness. Citing several studies (e.g., Davis 1968 and 1972, Spearitt 1972), Nation and Coady suggest that memory of word meanings is the only subskill consistent across various analyses in measurements of the text readability. Indeed, vocabulary knowledge is the most clearly identifiable subcomponent of the ability to read. Yap (1979: 58 in Nation and Coady 1988) states that "causal links probably do exist between vocabulary and comprehension and that vocabulary is likely to be the predominant causal factor."

The readability of a text, therefore, relies heavily on the type of vocabulary of which it is comprised. However, reading in second language is hampered by at least four other possibly related factors (Alderson 1984 and as cited in Laufer 1992):

1. poor reading ability in the L1;
2. inadequate knowledge of the L2;
3. incorrect strategies for reading in the L2;
4. reading strategies of the L1 not being employed in the L2.

To these four factors we might also add:

5. lack of knowledge of the text's topic;
6. lack of interest in the text;
7. lack of enjoyment of (or interest in) reading;
8. lack of enjoyment of a specific genre;

9. lack of enjoyment of a specific reading medium.

In addition to these factors, there is strong evidence that there is a threshold language competence level that must be attained before reading skills in the first language are transferred to the second language (Alderson 1984). However, we again see that lexis is at the centre of this competence level. Laufer (1984) examined three aspects of this threshold: (1) how vocabulary size correlates to reading comprehension; (2) what vocabulary size is needed to attain certain scores on comprehension tests; and (3) how an increase in vocabulary size affects comprehension scores. Laufer used standardised tests for reading comprehension (the reading component of a Dutch examination) and vocabulary size (e.g., *The Eurocentres Vocabulary Test* - Meara and Jones 1989). She found significant correlations between reading and vocabulary scores. These results confirm the notion that vocabulary is a good predictor of reading comprehension level (although not necessarily a causal factor), and subsequently strengthen the argument for a heightened concern with vocabulary.

2.2.5.2 The Relationship between Vocabulary and Grammar

There is some evidence that much of the information acquired about grammatical structures is lexical (Bland 1987, cited in Sonaiya 1991). The link between lexis and structure, therefore, is probably stronger than previously imagined. Sonaiya (1991: 275) writes that,

> ...the importance of studying issues relating to lexis within language learning needs to be highlighted. On the one hand, it is questionable whether the traditional division of language into grammar and lexicon is one that is valid....Furthermore, there

> seems to be evidence that some aspects of language that have hitherto been dealt with under grammar, particularly in the area of language acquisition, are actually lexical in nature.

Some researchers argue that word recognition abilities are the single best class of discriminators between good and weak readers (Adams and Huggins 1985). After all, 'knowing' a word entails knowing how it can be used syntactically. One can not underestimate the closeness of the relationship between syntax and syntactic knowledge of lexical items. It is not within the scope of this thesis to re-examine the distinction between lexis and grammar. Nevertheless, in light of the above arguments, the traditional dichotomous view is becoming increasingly unsatisfactory.

2.2.6 The Mental Lexicon

2.2.6.1 The State of the Mental Lexicon

L1 studies in the mental lexicon were popularised in a study by Fay and Cutler (1977). Their research, based on malapropisms, suggests that words are arranged phonologically but accessed by two different networks: one phonological and one semantic. In their study, ninety-eight percent of native speaker malapropisms had the same stress patterns and eighty-seven percent had the same number of syllables as the speaker's suspected target words. While a high percentage came from the same word class as the target word, form did not seem to play as important a role as stress, rhythm and phonology. The authors conclude that words are arranged in the mind semantically, phonologically, and structurally.

Aitchison (1987) writes that words can generally be divided into *content* (or *lexical*) words and *grammatical* words. The study of the mental lexicon is generally concerned with lexical words and their organisation. In the mental lexicon, a word's meaning and its part of speech are not separate entities but are integrated within the same component. Words are probably more tightly connected to other words in the same word class than even to the same form of a word that might appear in a different word class. For example, the verb *shout* is probably stored more closely to the verbs *yell* and *bellow* than to the noun *shout*. Proof of this appears in so-called 'tip-of-the-tongue' errors where the majority (80 %) of incorrect words come from the same word class as the intended word (Deese 1965 in Aitchison 1987). Guesses and word association experiments suggest that words from the same classes are quite closely linked, especially nouns.

Aitchison cites one of the primary questions of lexical research: is the mental lexicon stored as single items ready for use or are they disassembled into morphemes and then put together when needed? (Flores d'Arcais & Meijers 1983 in Aitchison 1987 and Cutler 1983). Aitchison looks first at derivations and inflections. Derivations are those attachments which when added form a new word (such as *teach* and *teacher*). Inflections are attachments that add information to a word without fundamentally altering it (such as *dog* and *dogs*). Evidence suggests that inflectional endings are added on when needed; that is, for example, *dog* is not stored as a separate item from *dogs* (Cutler 1983). Errors made by speakers confirm this (e.g., 'He *go backs*'; and 'They *point outed*'; Aitchison 1987: 109). One must be careful how strongly one argues that inflections are produced and not stored with the word as a separate item. Slips of the tongue occur with regular items (Aitchison gives the example of *Fran Sancisco*) suggesting that inflectional errors probably occur just before the time of utterance rather than in the planning stage preceding it.

Derivations, in the case of prefixes, are generally attached to stems and stored that way—at least with the more common words. This is due primarily to the fact that the rules for forming stems plus derivations are so erratic. There are, therefore, three reasons for making derivational errors (Aitchison 1987: 115):

1. Blending: two words are blended together as in 'it is fun to speculise' (speculate+surmise)
2. Derailment: the speaker starts a word, but not paying attention, gets carried onto another word as in, 'Bees are very industrial' (industrious).
3. Fall-Back Procedure: When people cannot think of a word, they use rules to make up a new word (which might actually be correct) as in, 'Children are so deduceful' (deductive).

It appears that there is a great deal of variety in the way words are accessed and that generalisations must be made with considerable qualification. Cutler (1983 citing Fay 1980) reports that prefixed words, for example, could be accessed holistically, but that access via the stem word was also possible. Thus, the complexity of the system often defies tidy distinctions between access processes.

2.2.6.2 Lexical Difficulty

Words vary in the difficulty they present to the second language learner. Cutler (1983) argues that words vary in complexity on three levels: (1) semantic, (2) syntactic, and (3) morphological. Semantically, words can carry different kinds of ambiguity. They can be unsystematically ambiguous (e.g., *bear*—has no semantic relation to an alternate meaning), they can be systematically ambiguous (e.g., *glue*—has a related sense in a different category), or they can have a related sense but distinct referents (e.g., *run*—the meaning is carried from the central meaning to other meanings with words like *river*, *car*, etc.). It is also argued that words are more difficult if they are used in a way that differs from their central meaning. Cutler goes on to say, however, that mental access time is not hampered by lexical complexity in first language use. It appears that lexically complex words are not more difficult to retrieve from the mental lexicon than lexically simple words. Cutler (1983: 73) also notes, however, that,

> many types of lexical complexity can lead to greater processing difficulty just in the case that (a) the task taps the time to construct a representation of the active sentence and/or to make judgements about it, *and* (b) the effect of the lexically complex word is to enable more than one representation of the sentence to be acceptable.

It is not clear whether access times to lexical items in an L2 are similarly unaffected by lexical complexity. However, one would expect that the problems L1 speakers have with constructing sentence level representations would be exacerbated in the L2 learner—to the extent that learners were aware of alternate or ambiguous meanings.

In an L2 learner, there are different factors influencing both the accessibility of a lexical item and his or her ability to understand it in context. Nation (1982) observes that factors influencing the difficulty or complexity of a lexical item in the L2 include:

- pronouncability;
- the item's part of speech;
- the item's similarity to words in the learner's L1;
- the learner's proficiency level;
- the kind of test used to gauge learners' vocabulary knowledge (Nation 1982).

Laufer (1986b) adds that length of the word, the inflexional complexity of the word, its derivational complexity, its abstractness or concreteness, its idiomaticity, and its register restrictions can work together wholly or partially to increase lexical difficulty. In non-related languages, words in the target language may conflict with 'borrowed' words in the L1. For example, similar to the 'false friends' in many European languages, Japanese (which distinguishes borrowed words by using a differently shaped syllabary) has many words that native speakers of Japanese might consider to be 'English' but which may, in fact have either lost the original English meaning (e.g., *service* means *no charge* in Japanese) or might have originated from a different language altogether (e.g., *arbeit*—German—means *part time job* in Japanese). With all these factors in mind we can safely argue that lexical complexity is likely a great source of learner error. Furthermore, what we know about lexical difficulty makes discussions about what students learn in an experimental situation more difficult. The observed effects from a given experiment could be as much a function of the kind of words (in terms of their intrinsic difficulty) learners encounter as it is a function of the particular method in question.

2.2.6.3 Sounds and Sound Patterns

Aitchison also discusses the significance of sound patterns in storing the mental lexicon. She writes, "A word can be likened to a body: flesh (the sounds) covers an underlying skeleton (the rhythmic pattern) which gives it its shape." (Aitchison 1987: 121). In spite of minor differences, all researchers agree on one thing: words have a rhythm which is likely to be specified in the mental lexicon. The important point about this rhythm is that it is relative. The exact amount of stress given to each syllable may vary (in English), but it is essential to get the relationship of strong to weak right along the whole length of the word (Aitchison). The importance of stress, rhythm, and sound in remembering words cannot be overemphasised. All these acoustic factors assist in the arrangement of words in the mind and the recall of a word at the appropriate time. Indeed, neurological studies in aphasics suggest that acoustic information provides at least half the information the brain needs to process lexical information (the other half is provided by orthographic information (Coltheart 1987)).

The model proposed by Fay and Cutler, Aitchison, and others is consistent with the view that syllable recognition and stress recognition play important roles in the mental lexicon. Hence, to optimise comprehension of such languages, methods used to present new vocabulary to learners should lead them to internalise accurately pronunciation of individual sounds, number of syllables and stress (Channell 1988).

2.2.6.4 Implications for the L2 Mental Lexicon

Channell (1988) suggests three questions that should be asked pertaining to second language lexical research: (1) how much does the mental lexicon of the L1 resemble that of the L2 in any learner, (2) is the bilingual's linguistic mental organisation similar to the monolingual's or qualitatively different i.e., is bilingualism more of the same or something completely different?, and (3) what is the resemblance, if any, between the L2 learner's lexicon of a language, and the mental lexicon of the native speaker's lexicon of that language? Channell argues that in the absence of evidence to the contrary, it is acceptable practice to draw on L1 theory to test hypotheses in L2 theory.

Quoting Tarone (1974) Channell suggests that in the beginning stages of learning to comprehend, learners rely on selective processing (they have difficulty coping with everything). Learners use stress to find out what is important. Hence "for comprehension, learners need to know both the citation form stress of a word, and also the permissible permutations of stress placement that a word can undergo in continuous speech." (Channell 1988: 90). Channell makes the pedagogical point that teachers often use unnatural stress to 'help the students to understand'. Students receiving such instruction often learn that unnatural stress and are confused when speaking outside the classroom (when their interlocutors do not seem to understand, or when they do not understand their interlocutors).

As Fay and Cutler found in their L1 research, syntax and grammar undoubtedly play important roles in L2 learners' storage of mental vocabulary. However, the exact nature of that role has not yet been researched extensively. Learners do, nevertheless, use associations between phonology of the L1 and new L2 words. This is done primarily for the sake of memory. For example, a new word may sound like a phrase in the L1; learners use mental 'tricks' to remember words. Research in this area gave rise to the 'key word' method of vocabulary learning (learners associated new words with similar sounding English words; mental associations were then made between the two). Cohen and Hosenfield (1981), for example, found improvements in vocabulary retention when students were instructed to make associations between L1 and L2 words. Researchers have also found that mnemonic devices (like associating words to bizarre pictures) increases performance on vocabulary retention tests (Ott, Blake & Ball 1974). Studies with learning impaired students (and later with normal students) revealed that learners who made use of mnemonic strategies showed strikingly fewer errors in pictorial association tasks (Ott, Blake, & Butler 1976, citing Jensen and Rohwer 1965). The primary difficulty with paired associate learning, however, is that it ignores the "complex patterns of meaning relationships that characterise a proper, fully formed lexicon as opposed to a mere word list." (Meara 1983: 104)

Carter and McCarthy (1988 :12) confirm the importance of associative devices in learning and retaining vocabulary:

> The principle of vocabulary learning which emerges is that the more words are analysed or are enriched by imagistic and other associations, the more likely it is that they will be retained.

The richer the context then, the more likely a student will remember both the word and the context in which it was encountered. One would expect that if a learner both read and heard a word, s/he would be more likely retain that word for a substantial period of time.

Channell makes three conclusions concerning the state of the L2 mental lexicon.

1. For L1 there is one mental lexicon, phonologically arranged with word stress, syllable structure, and syntax acting as high level organisers. The lexicon is accessed by distinct but interrelated networks for production and perception;
2. L1 and L2 lexicons within the same speaker are clearly linked, phonologically, semantically, and by association;

3. Evidence that L1 lexicons and L2 lexicons of the same language resemble each other is sparse.

Henning (1974) made a more direct examination of the L2 mental lexicon. He investigated two elements of mental vocabulary arrangement in L2 learners: (1) semantic clustering and (2) acoustic clustering. The author tested two hypotheses:

1. That second language learners will make significantly more errors on a vocabulary recognition test by choosing semantically and acoustically associated distractors than errors by choosing distractors which bear no semantic or acoustic association with the correct recognition responses;
2. That learners with higher proficiency levels will use semantic encoding more that learners with low proficiency levels and that higher level learners will use less acoustic encoding than lower level learners.

Since first language studies show that children start with a high degree of acoustic association and then acquire progressively more semantic associations (Bach and Underwood 1970 cited in Henning 1974) one might expect that L2 learning follows the same pattern. If this can be established, educators would better be able to design appropriate materials for learning vocabulary by emphasising the acoustic nature of words as well as their orthographic representations.

Henning tested learners by exposing students to aurally presented text. Words were randomly chosen from the text and given distractors that were either semantically related, acoustically related, or neither semantically or acoustically related. The author found that acoustic mistakes were more common at lower levels and semantic mistakes more common at higher proficiency levels. He writes that,

> ...second language learners do encode vocabulary in short-term memory in clusters according to associations in meaning and sound. Furthermore results...confirmed the hypothesis that short-

> term memory encoding by acoustic clusters is negatively correlated with language proficiency (Henning 1974: 192).

Thus the level of the student may determine the amount of acoustic information used. We would not expect advanced learners to disregard acoustic information completely but instead to use semantic information which is not available to lower level learners. One might argue that acoustic supplements to the reading process (as, in this case, the computer is proposed to be) are more suitable for lower level learners who are learning to distinguish sounds and match sounds to orthographic patterns (especially in learners with non-alphabetic first languages) than to higher level learners who use semantic information predominantly. It would be difficult to imagine, however, that acoustic information would decrease performance in higher level students—unless the sound was somehow distracting (too loud, poor quality, in an irritating voice, etc.).

While it cannot be argued that the L2 mental lexicon follows all the patterns of the L1 lexicon, clearly, acoustic factors play an important (if not completely understood) role in both the L1 and L2 lexical organisation and comprehension. The L2 learner who is not as well-equipped as the L1 child in knowing the importance of a word through its structural placement, uses stress to determine a word's relative importance. In the present study, we might argue that the presentation of text in both a written and aural context should encourage the internalisation of words by providing additional mental references (i.e., memory of the sound as well as the text). From the student's perspective, hearing the stress patterns should make the relative importance of words and sentences more accessible.

It is a central issue in this thesis that the addition of sound for mental reference to context, and the provision of stress as a signal of the relative importance of words, will produce marked improvements in vocabulary acquisition. This assumes, however, that students are aware of the role of stress in English and that they are capable of taking advantage of the additional information—despite the possible unfamiliarity of computers as the instructional medium.

2.2.7 Theoretical and Practical Issues

Examination of different methods for learning new words in a second language have existed for some time. Forlano and Hoffman (1937), for example, examined the differences in acquisition of words where one group of students guessed word meanings and the other group were told the meanings by the teacher. The 'guessing' group were given no clues to the meaning of the word, the 'telling' group were told what it meant. Not suprisingly, the 'telling' group performed better on the delayed retention test than did the 'guessing' group. Research such as this resulted in the dubious pedagogical generalisation that it is better to tell the correct meaning of words to students immediately without allowing the learner to guess a possible wrong meaning.

In a more current context, Richards (1976) argues that theoretical researchers do not necessarily have teaching the language as their goal. Instead, they are most likely to be concerned with the nature of the language and descriptions of how language is used to carry out pragmatic functions. Richards (1976: 77) succinctly states that,

> Inevitably such information [information gained from theory-centred hypotheses] will turn out to be vastly more complex than we might intuitively have supposed, yet will be tentative and inconclusive because of the changing state of knowledge and theory in the disciplines concerned. Such information cannot be translated directly into teaching procedures.

Theory that wholly supports one method in exclusion of another risks the scepticism of teachers who may very well 'know' that their method works quite well—regardless of the purpoted evidence that may support the theory.

In the area of vocabulary instruction and acquisition, the theory/practice discussion has partially manifested itself by the favouring of active teaching over passive vocabulary learning. On the 'active' side of the debate, teachers have observed that students discover the meanings of new words through various means: (1) asking others, (2) using a dictionary, and (3) being told to make use of context and to guess from the item itself (e.g., Gairns and Redman 1986; Oxford and Scarcella 1994). On the 'passive' side of the discussion, some theorists argue that vocabulary can be 'taught' by not actively doing anything: just expose the learners to lots of words in an intelligible context and they avoid 'teaching' (i.e., drilling) what they consider to be, ' the hard way' (Krashen 1989). Vocabulary in this view is acquired incidentally. It may very well be true that vocabulary can be learned (or 'acquired' as proponents of this view would most certainly prefer) without instruction. There are two dangers to an extreme view, however. First it is contrary to what most teachers believe, observe, and practise in their classrooms (cf. Lightbown and Pienemann 1993). Second, it goes against what is known about learner differences (different learners learn in different ways) and the variable nature of second language acquisition and interlanguage (generalities and universals may exist but are difficult to pin down). Third, by de-emphasising the active ways in which vocabulary can be taught, we risk vocabulary falling again into 'poor relation' status (Carter 1987).

The argument in this thesis that words can be learned incidentally from a reading context does not exclude other sources of vocabulary learning. For example, in second language studies there is a great deal of research suggesting that the use of dictionaries can be effective and that more investigation is needed into the best way of presenting word definitions and examples on-line or otherwise (e.g., Cumming, Cropp and Sussex 1994)—although dictionary definitions can be misleading for users and misinterpreted by adult second language learners (Nesi and Meara 1994). Also, as has been said, there is a strong argument that words are better learned in pairs than in context at certain stages of language development (e.g., Nation 1982). Vocabulary can therefore appear in authentic contexts (short or long), long teacher-contrived contexts or very short contrived contexts; words can be learned in lists, with visual imagery, by 'physical response' or any number of other 'teaching' means (Oxford and Scarcella 1994).

Nevertheless, the prospect that students learn vocabulary from reading is an appealing one (also assuming that reading *per se* is generally considered desirable for other reasons). If incidental acquisition is one of many means of acquiring vocabulary, then it is worth investigating creative methods of increasing the students' likelihood of acquiring vocabulary incidentally through reading. This might be done by enhancing the reading materials themselves and/or by making the materials more appealing so that students are encouraged to read more.

2.3 Incidental Vocabulary Acquisition

Reading in a child's first language is generally seen as a source of a large amount of vocabulary (Nagy, Herman, and Anderson 1987). Extensive reading is also regarded by many to be a good source of vocabulary in second language learning (Krashen 1989). Research with languages other than English is also beginning to show that in children, a large number of words are acquired through reading (e.g., Shu, Anderson, and Zhang 1995). Oxford and Scarcella (1994: 240) write that,

> We believe that the learning of words through written and oral contexts constitutes a major factor contributing to L2 vocabulary development....Regular, sustained reading, in particular seems to be an effective way to promote large-scale L2 vocabulary growth.

Sophisticated methods to determine how vocabulary is acquired by L2 learners are constantly being developed. Generally speaking, research in this area seeks to expose students to new vocabulary in some sort of context and then measures increases in average scores on a post-test. Any gains in understanding of unknown words can then be attributed to the experimental treatment. Such research usually includes a control group who are not exposed to the new words in any way and an experimental group who typically do a reading activity or combined reading/topic-familiarisation procedure. The measurement tool might take the form of a multiple choice test with the new word as the test item and a series of definitions as the test answers and distractors. The following section will describe some key studies in first and second language acquisition.

2.3.1 Reading

2.3.1.1 First Language Incidental Vocabulary Acquisition

The terms 'incidental learning' and 'incidental acquisition' are used extensively throughout this research. In the psychology literature, incidental learning is often referred to as 'implicit learning'. As such, Reber, Walkenfield and Hernstadt (1991: 888) have defined it as:

> the process whereby a complex, rule-governed knowledge base is acquired largely independently of awareness of both the process and the product of the acquisition.

It is thought, and the Reber *et al.* study supports this notion, that the capacity to learn implicitly does not vary greatly among individuals. This is contrasted with 'explicit learning' which is thought to be highly variable between learners and strongly correlated with cognitive ability as demonstrated by IQ tests (Reber, *et al.* 1991).

It is important to understand the context of literature when discussing incidental acquistion, However, this present research is not designed to determine differences between implicit and explicit learning, nor are we attempting to assess the relationship between implicit and explicit learning in terms of learners' cognitive ability (as Reber *et al.* did). Like Reber *et al.*, students are not given any instructions other than to read the text and how to manipulate the interface. Vocabulary acquisition is assumed to be incidental because students have no external motivation to apply explicit learning strategies to newly encountered vocabulary. In fact, studies with learning strategies (see Section 5.1.2.3) show that good students actively attempt to decipher and remember new words. Indeed, it is hoped that students will do this in the present experiment, because it will make any differences between the listeners and the readers more apparent. Precedents for this view of incidental vocabulary learning can be seen in Krashen (1989), Day, Omura, and Hiramatsu (1991), Hulstijn (1992), and Shu, Anderson, and Zhang (1995). By way of an operational definition then, incidental acquisition means the learning of new (or the clarification of unfamiliar) words from a reading or a reading-while-listening context, without any suggestion by the teacher as to how or if that learning should take place.

Jenkins, Stein and Wysocki (1984), examined the hypothesis that students could learn the meanings of unfamiliar words from context when they are given relatively normal classroom reading tasks, i.e., tasks that do not call extraordinary attention to vocabulary. The authors presented a group of fifth graders (eleven-year-olds) with low frequency words in enriched contexts (by enriched, we mean, the words appeared several times and in contexts where the meaning could be easily inferred). Results from this study indicate that increased encounters with new words in context resulted in greater word learning and more than two encounters were needed to affect vocabulary acquisition. At the same time, however, the authors (p. 784) note that,

> The most intriguing finding in the present investigation was not that some word meanings were learned during reading, but that the

> number of words learned was fewer than might have been anticipated.

The authors note that reading itself is not a satisfactory explanation for vocabulary growth. One clarification should be made, however. The authors in this study were primarily searching for an explanation for the large amount of vocabulary learning that takes place in a child's early schooling. They were not investigating enhancements to vocabulary acquisition or examining the effects of a new method. If they had, the results from their study would have been a satisfactory justification for more reading despite it being an unsatisfactory explanation of vocabulary development.

2.3.1.2 Second Language Studies

Hulstijn (1992: 122) argues that it is a plain but important fact that

> ...the retention of word meanings in a true incidental learning task is very low indeed. The chance that L2 learners/readers remember the meaning of a word, occurring only once in a text which is being read for its content, is very small.

He does not claim that incidental acquisition does not take place, however. Indeed, the consensus among researchers is that incidental acquisition through reading occurs, but in very small, sometimes negligible increments (Hulstijn 1992). Similarly, Nation and Coady (1988), observing that studies of learning words from context have not shown the large amount of learning that might be expected, attempted to investigate small incremental changes in vocabulary after reading. They estimate that the chances of acquiring a word after one exposure is between .10 and .15. Saragi, Nation and Meister (1978) have shown significant correlations between the number of times an unknown word appears in a book (in this case, it was *A Clockwork Orange* and the words were *nadsat* words) and the number of people who chose the correct meaning in a post-test. Saragi *et al.* (1978: 76) concluded that "a considerable number of repeated words can be learned incidentally through extensive reading by meeting them in context without reference to a dictionary."

However, recent studies (e.g., Pitts, White, and Krashen 1989) have also demonstrated significant gains in short term retention of vocabulary after brief exposure. Pitts, *et al.* (1989) set out to determine what effects reading has on new vocabulary acquisition. Pitts *et al.* used *A Clockwork Orange* which contains words created by the author (*nadsat* words). An experimental group read the book and were thus exposed to the pseudo-words in context. The control group did not see the nadsat words before the test. As might be expected, the control group showed almost zero knowledge of the new words. The experimental group, however, showed small but significant increases in knowledge of the new words.

Following the example of this study, Day, Omura and Hiramatsu (1991) hypothesised that Japanese EFL students could learn vocabulary by reading silently for entertainment in the classroom. It is worth looking at this study in some detail because to a large degree, the method in this thesis follows the pattern established by Day *et al.*. The authors used Japanese High School and university EFL students and pre-tested 27 target words in a pilot study. On the basis of this study, 17 words were chosen and written into a story of about 1000 words. The story was re-written so that the words appeared frequently and in ample contexts throughout the text. Students were assigned randomly to either a control or experimental group. The experimental group were given the story to read then give a multiple choice post-test; the control group was simply given the test. The test items included the correct answer plus three distractors and an "I don't know" option. The individual student scores were corrected for guessing using the following formula:

Score = right - (wrongs/n-1), where n is the number of choices.

On an independent samples t-test, Day *et al.* found significant differences between control and treatment group post-test means ($p < .01$). They conclude (p. 545) that,

> The finding of a causal relationship between reading and indirect vocabulary learning in an EFL context is consistent with research into vocabulary learning by children in their first languages....The investigation...provides empirical evidence for the claim that foreign language students can learn target vocabulary through reading.

Based on the evidence from this study, we suggest that in certain experimental situations, small increases in vocabulary learning will occur after brief exposure to new words in a written context. In the case of incidental acquisition through reading, for example, experimental results indicating small increases in vocabulary acquisition after one reading probably indicate that extensive reading will result in greater vocabulary acquisition (Day *et al.* 1991).

2.3.1.3 Related Issues

Other research has recently focused on what cues or contexts are most likely to assist learners in the acquisition of vocabulary while reading. Hulstijn (1992), for example, has examined several different ways to encourage learners to think about the context as they are reading, thereby enhancing the incidental acquisition of words. Hulstijn and others with similar interests base their experiments on two principles (also known as the *mental effort hypothesis* (Hulstijn 1992: 113)):

1. When subjects have to infer or induce the solution of a problem, they will invest more mental effort than when they are given the solution to the problem;
2. Information that has been attained with more mental effort can later be better retrieved and recalled than information that has been attained with less mental effort.

To investigate these hypotheses, Hulstijn constructed an experimental design where subjects in different groups were given a reading comprehension task. The text consisted of mostly high-frequency words interspersed with lower-frequency or pseudo-words. The subjects were either given the meaning of the words in the margin, given cues (in the form of multiple choice alternatives, or a concise context using the target word), or no information was given at all. The subjects did not know that they would be post-tested (the researchers did not want the students to use any extraordinary techniques for remembering/learning the new words). In the post-test, Hulstijn found that increased mental effort (through inferring word meaning by themselves as opposed to being given the word meaning) increased the likelihood of the learner remembering the form and meaning of the unknown word. Hulstijn argues that researchers should not debate whether or not to give word meaning, but which cue is the most effective means of increasing mental effort.

The effects of mental effort on retention of L1 words are well established. Craik and Tulving (1975) found that first language learners remembered word pairs better if they were asked to use more processing skills to arrive at a solution. The processing and problems that the subjects were asked work through were not difficult—often no more than filling in blanks in the middle of words to determine the word in question. Craik and Tulving found that subjects "remembered not what was 'out there' but what they did during the encoding." (Craik and Tulving 1975: 292).

Even though we are not arguing that learners' mental effort is increased by the computer-based materials described in the next chapter, it is apparent that hypertextual links could be used very effectively using Hulstijn's experiment as a model. For example, multiple choice alternatives for target words could appear in the margins on screen and students could have immediate access to correct answers, hints, and explanations. Notions of mental effort should be kept in mind, therefore, when the question of additional hypermedia features arises in Chapter 5.

Most of the research in incidental vocabulary acquisition supports the view that lexical meaning is better retained if learners are forced to expend some effort or if the context is enhanced or 'rich'. The following section will examine other means of enhancing a reading context and how the computer could be used to do so.

2.3.2 Listening while Reading

This thesis considers the addition of a spoken text to a reading exercises as one possible enhancement to encourage incidental vocabulary acquisition. Specifically, it questions whether the small gains in incidental vocabulary acquisition seen in reading could be more noticeable by adding an aural context to the written text. Relatively few studies have investigated this to date. This may be due to the fact that reading aloud extensively to students is not done by most teachers. Therefore, there may have been little point to studying its effects if teachers could not envision themselves (or researchers could not envision teachers) reading aloud to students. Despite the lack of research in incidental vocabulary acquisition from reading and listening, the possibility of using computers for the task has resulted in preliminary research investigating its effects in first language learning. As with many aspects of language theory, studies in first language acquisition have preceded investigation in second language research. The following section will look at some first language investigations and that will be followed by an examination of the literature in second language research. This discussion will not be limited to vocabulary acquisition but will also include the benefits of reading while listening for comprehension and attitudes towards reading.

2.3.2.1 First Language Studies and Reading while Listening

Since it is generally acknowledged that reading aloud to children develops their reading and comprehension skills (cf. Chomsky 1990), studies are beginning to examine the effects of listening while reading on vocabulary acquisition. Children like to be read to. There are, however, obvious psychological variables when reading to children (e.g., increased personal attention) that may or may not apply to L2 vocabulary learning. Nevertheless, interest in reading aloud and reading with sound has led to increased research in L2 experiments with both teacher initiated and computer initiated voice input.

Reitsma (1988), for example, conducted experiments into the effects of computer-generated sound on vocabulary acquisition. Interestingly, the footnotes describe a 'future' system that would "allow for presentation of text on a screen accompanied by highly intelligible speech..." (p.235). Affordable systems are now available to do exactly what Reitsma envisioned. He compared the effects of guided reading, independent reading, and self-selected computer-generated speech feedback on new-word comprehension of first language readers. The author's results can be summarised as follows (Reitsma 1988: 219)

> Both guided reading and independent reading with self-selected speech feedback were found to be significantly more effective than the control and reading-while-listening conditions. The findings suggest that increases in reading efficiency depend largely on the amount of independent, self-propelled reading activity of young readers. If such independent activity is included, computer-aided practice with speech feedback seems promising as a means of improving reading skills in beginners.

Rasinski (1990) tested the effects on first language readers of listening repeatedly to texts and listening while reading. There were two major findings:

1. Both repeated readings and listening-while-reading treatments were effective in improving the reading fluency of third-grade students;
2. Neither treatment was superior to the other in improving students' reading fluency.

Elley (1989), using a control group/treatment group design found that reading aloud to children is a significant source of L1 lexical acquisition. Furthermore, if the teacher provides an explanation of new words as they are read, vocabulary gains can more than double. The new words are also relatively permanent. Elley used a delayed post-test to measure the retention of new vocabulary and found that decline in performance between mean scores of the immediate post-test and the delayed post-test was a negligible 2-3 percent.

We can infer then, that extensive reading and reading while listening (with the possible option of self-selected feedback) are viable means of increasing L1 vocabulary. The question remains to what degree L2 teaching can assume the same.

2.3.2.2 Second Language Studies and Reading while Listening

Before looking specifically at vocabulary acquisition, and because this thesis is also concerned with affective and comprehension factors, it may be a good idea to examine some other benefits of reading while listening in terms of attitude and comprehension. In a Middle Eastern context, Dhaif (1990) questions the currently popular notion that reading aloud to students is not as good as solitary silent reading. Dhaif's reviews the research of childhood reading development in an L1 and how reading aloud assists their comprehension. It is Dhaif's contention that reading aloud is a valid yet neglected method of assisting L2 reading comprehension. His research supports this hypothesis and also suggests that students like their teachers to read aloud to them (at least in this particular cultural context): Subjects read in two sessions. In the first session, they read silently, alone, and in the second, they read along as the teacher reads aloud.

Dhaif cites two American doctoral dissertations that may be relevant in this part of the study. Summarising May (1986) he says, that current research shows substantive evidence indicating the value of reading aloud to children. He also notes, however, that there is little similar evidence supporting the use of reading aloud with non-native speakers (Dhaif 1990).

Dhaif (1990: 458) similarly acknowledges Santos' (1987) conclusions:

> The results of this study revealed that reading aloud has a significantly positive effect of the ESL learner's comprehension skill, particularly in their ability to inter-relate, interpret, and draw conclusions from the content. Both studies suggest that further research is needed to investigate the effect of the reading aloud technique on adult learners and learners from other cultures.

Dhaif tested students' attitudes towards reading along with the teacher, and he compared comprehension scores between silent solitary reading and reading as the teacher read to the group. Qualitatively, Dhaif found that 77% of the students (N=140) preferred the teacher's reading aloud technique and among these, 47% indicated that it helped them understand the overall meaning of the text better. Quantitatively, Dhaif found a statistically significant difference between the test scores of the two reading sessions, suggesting that for learners of a similar level (here the author says his students have a "basic level of proficiency in English..." p.463) and background, reading with a teacher who is reading aloud improves comprehension. Furthermore, the author suggests that reading aloud will increase comprehension to the degree that readers will be freed from the "vicious circle in which weak readers tend to be trapped: they do not enjoy reading because they do not understand, and because they do not understand, they do not read." (Dhaif 1990: 463). The results from this study must be seen in terms of their specific cultural context. In countries like Japan, for example, reading aloud is generally considered an indication of the teacher's inability to use English in any other productive way

(Inoue 1994). On the other hand, it may well be that students in other countries consider the reading aloud exercise as an indication of the teacher's personal interest.

2.3.3 Reading while Listening: Time Compressed and Time-expanded Speech

The work of A.K. Pugh and M.H. Neville should be central to any discussion of simultaneous reading and listening. Neville and Pugh (1973), using machine-assisted texts, examined the effects of reading while listening on comprehension, reading rate, and reading fluency. Despite the fact that hypermedia tools were not available to them, they were able to utilise a system wherein reading and listening materials could be 'time-compressed' or 'time-expanded', i.e., recorded texts could be sped up or slowed down through a mechanical device. One of the problems with tape recorders is that to manipulate the speed of the oral text, one has to tolerate significant deterioration of sound quality. Neville and Pugh, however, managed to reduce temporal space within and between words without any significant loss in acoustic quality.

In addition to describing this mechanical innovation, Neville and Pugh argue, based on the work of Sokolov (1960), Conrad (1972), Edfelt (1959), Anderson and Dearborn (1952), and Neville (1968), that 'listening to an oral reading by an accomplished reader while following the text silently, improves the fluency of beginning readers...' (Neville and Pugh (1973: 2)). A parallel argument to the notion that simultaneous reading and listening improves fluency is the idea that 'pacing' the text will improve comprehension. Citing Foulke and Sticht (1969), Neville and Pugh write that using recorded speech as a pacer will allow for more processing time, making difficult texts more comprehensible and also, that if the pace of texts could be increased so that the reading was faster than normal speech, the level of difficulty could be altered without changing the complexity of the written text itself.

Neville and Pugh, sought to determine whether pacing of reading through the slowing of speech for silent reading would improve the understanding of the passages and whether the pace could be increased without loss of understanding. In keeping with the notion that reduced subvocalisation is desirable and that paced oral supports reduced subvocalisation, the authors hypothesised that subjects would show a decline of observable subvocalisation while showing improvement of their English reading ability.

The authors found that expanding the text (i.e., pacing the text so that it takes *more* time to read), apparently does make the text easier for non-native readers of English and that compressing the text makes it more difficult to comprehend. On the other hand, the authors note that pacing the text may make it generally more difficult than silent reading because the readers are forced through the text without being able to look back and reprocess the reading more carefully (when they encounter difficulties). The authors' results also seem to confirm the hypothesis that subvocalisation decreases over time with paced reading.

2.3.3.1 Implications

The implications of Neville and Pugh's work are central to several assumptions in this thesis and interact with our arguments here in at least three distinct ways: (1) in terms of the mechanics of materials development, we assume that computers are efficient and relatively easy tools for developing or researching reading and listening materials. Computers should, in practice, be easier to implement than a system such as Neville and Pugh's ; (2) in theoretical terms, their research implies that pacing the reader through the use of sound may have a beneficial or deleterious effect, depending on the speed of the pacing; and (3) pedagogically speaking, materials that make use of simultaneous reading and listening should be at least as desirable and practicable in a classroom context as Neville and Pugh's system. We will examine each of these more specifically in turn.

First, on the mechanical level, the effect of the pacing machine used by the researchers (the Eltro Tempophon Speech Compression Machine (Neville and Pugh 1973)), should more readily obtainable with computer-based systems. Even if automated algorythms are not available for compressing or expanding an aural text, software is available that would make 'manual' (through on-screen editing) compression/expansion relatively easy. The length of individual words and spaces between words could be manipulated using standard bundled software that comes with most Windows-based sound systems. For the most part, this manipulation could be done without significant negative affect on sound quality. Certainly, it could be done without changing the pitch of the text (which is what Neville and Pugh claimed their machine was capable of). For the purposes of experimentation, the present study could, in a future implementation, be adapted to include pacing at the word level. The premise that hypermedia software is appropriate for reading while listening, paced reading while listening, and reading, is therefore substantiated by the work of Neville and Pugh.

Second, on the theoretical level, pacing a text should be at least as effective (in terms of comprehension, reduced subvocalisation, and fluency) on a computer-based system as it was in Neville and Pugh's pre-PC mechanical system. In this regard, the present experimental system differs from the pacing tested by Neville and Pugh. It was the intent of this researcher to permit students to choose their own reading speed themselves at the sentence level in order to preserve their option to re-read the text. This is achieved through restricting the sound segments to the length of the sentence. The reader is in control of how quickly they listen to the following sentence. Neville and Pugh's system differs significantly in this regard because students had no control over the pace of the reading at any level, sentence or otherwise. As the researchers themselves pointed out, this may increase the difficulty of paced reading over normal silent reading.

On the other hand, it is possible that the pace within each sentence recording should be controlled by the compression/expansion method, especially where experimental results with un-paced reading show little variation between groups who use or do not use sound-accompanied reading, or where there is no improvement in vocabulary acquisition, comprehension, or fluency over time. Unless there is something fundamentally different between machine-controlled word-level pacing and user-controlled sentence level pacing, there appears to be no *a priori* reason suggesting that we will not see a significant effect on vocabulary scores and comprehension. Nevertheless, it is possible that the positive effects discovered by Neville and Pugh resulted from the fact that students were forced to maintain a pace throughout the duration of the text. Neville and Pugh's introduction to their research substantiates this possibility in its discussion of compressed speech The authors noted that studies with children showed significant gains in vocabulary scores through paced reading and listening (e.g., Reilard 1970; Reid 1971). The implication is that the positive effects were a result of the pacing rather than the listening *per se*, although the authors do argue that the acoustic element plays a significant role in improving reading comprehension. We do ascertain a certain risk, therefore, in not pacing the reading beyond the sentence level, and in not controlling the reading from start to finish. If the pace is too fast for the subjects involved, we would expect a negative effect on comprehension similar to that found by Neville and Pugh; if the reading is not controlled by the computer, the improvements in reading accuracy and rate found by Foulkes and Sticht (1962) might not be achieved.

Finally, on the pedagogical level Neville and Pugh's results suggested that:

1. certain students enjoyed reading and listening and felt that the technique improved both their oral English and reading;
2. listening helped all the students increase their pace and comprehension;
3. students were able to read more fluently with improved comprehension after training;
4. the speed of listening can be gradually increased to improve students' reading rate.

If the method is pedagogically desirable, then the present study should show similar positive results in terms of enjoyment, perceptions of fluency, and comprehension (as well as positive effects on some specific aspect of language development). The desirability of any method is determined to a large degree by the expectations of students before they encounter it and impressions after they experience it. Again, the results from the experiment to follow will demonstrate just how desirable a reading and listening programme is for a classroom context. Moreover, the gradual speed increase advocated by Neville and Pugh seems to be a legitimate way of encouraging students to read faster.

2.3.4 Other Benefits

The use of computer-based materials as a kind of surrogate reading teacher has not been examined in terms of vocabulary acquisition. What is known about the acoustic nature of the mental lexicon suggests that in circumstances where students are not exposed to spoken as well as written English (as is the case in countries like Japan), listening while reading would provide learners with the orthographic/phonological link that might be lacking in their own learning situation—thus further developing the potential to use the acoustic elements of the mental lexicon retrieval process. Moreover learners can rely on factors like stress to show the relative importance of words/phrases/sentences and to acquire the 'rhythm' of the word (cf. Aitchison in section 2.2.6.3 above) and the acoustic information that low-level students use (cf. Henning in section 2.2.6.4 above). If students' concentration is engaged by listening (as well as by initiating the listening themselves with the interface device), then one might expect a greater understanding of new or vaguely understood vocabulary items resulting simply from increased concentration.

As has been said, words are usually 'known' with some degree of vagueness, and 'knowing' a word entails different kinds of semantic and syntactic knowledge, words that are only vaguely known might be more clearly understood if the context (now having both visual and acoustic information) is better remembered. Studies should therefore not be limited to examinations of completely unknown words. It is possible that the greatest effects might be found with medium frequency words (e.g., words that have been encountered once or twice before) rather than with low frequency words (where one encounter in a possibly non-rich context might not be enough for understanding) or with high frequency words (that are probably already known by learners anyway).

Clarke and Nation (1980) have discussed the ways in which learners guess the meanings of words from context. Despite the fact that complete information about a word is not always available from one or two contexts, there is quite a lot of information available the would allow students to decrease the vagueness of an unknown word. Among these are:

1. Knowing the part of speech of a word;
2. Knowing whether the word has a positive or negative connotation;
3. Inclusion words such as *and, in addition, furthermore*, etc.;
4. Exclusion words such as *or*, *not*, *alternatively*, *else*, etc.;
5. Cause and effect words (e.g., *because*, *as a result*), contrast words (e.g., *but*, *although*), conditionals (e.g., *if*, *provided that*), time references (e.g., *when*, *before*), arrangement (e.g., *firstly*, *finally*), classification (e.g., *comprises of*, *consists of*), comparison (e.g., *similarly*, *likewise*).

Other, contextual clues, include how often the word appears in a text, whether or not the meaning of a word is repeated or rephrased, whether the word is part of a description or being described, and the word's part of speech as it may be determined by its placement in a sentence (Haynes 1993). Of course, not all learners are equally skilled at using all the available contextual clues.

Rather than thinking only of 'acquisition' we should be concerned with reducing the vagueness of words. Studying the ability of learners to remember and understand words as they were used in a reading/listening context, therefore, should clarify our understanding of the strength and importance of acoustic information for vocabulary acquisition.

As research has shown, L2 learners have different problems from L1 learners when words are guessed from context (Huckins and Haynes 1993). Second language learners, for example, can misrecognise word forms. If L2 readers think they already know a word form, they will simply access that meaning rather than looking to the context for information. Furthermore, L2 learners lack the background knowledge that L1 leaners have—and this inhibits the L2 learner's ability to use the context (Huckins and Haynes 1993). These problems are not insurmountable. Indeed, the computer's ability to highlight text for different readers, gloss words contextually, and react with great speed might provide a partial solution to these problems.

2.4 The Computer as a Source of Incidental Vocabulary Acquisition

As has been said, there are many ways in which computer-based materials can be developed in an attempt to improve vocabulary acquisition and comprehension, while at the same time providing an enjoyable learning experience for students. This section will provide a transition between Chapters 1 and 2, and

Chapter 3. It will discuss recent research in the use of on-line glossaries as a basis for initiating the discussion of vocabulary acquisition through reading and listening.

2.4.1 On-line Glossaries

The use of on-line dictionaries was mentioned several times in Chapter 1 (Sections 1.2.1 and 1.2.6). Intuitively, it is one of the more obvious means of assisting reading—it is relatively easy to include in a computer document and the computer's speed makes it superior to traditional dictionaries. Preliminary research into student use of on-line dictionaries is beginning to take place. Hulstijn (1993), for example, examined the look-up patterns of different types of reader by logging their use of an on-line dictionary. In his sample of Dutch secondary school students, the subjects' tendency to look up words was not affected by their ability to infer word meanings in general, nor was it affected by their overall vocabulary knowledge (students were pre-tested for inferencing skill and vocabulary knowledge). He also found that the patterns found in the log data (number of times the text was read, etc.) was not related to performance on the comprehension post-test. Hulstijn found that a student's decision to look up the meaning of a word in a foreign language text is influenced by the perceived relevance of the word. Students were given a post-test to consult before and during the reading of the experimental text; then they were asked to take the test. The 'relevant' words, according to Hulstijn, were the words for which students knew they would be post-tested.

Hulstijn was primarily concerned with examining student decisions to use or ignore a dictionary. However, his research raises another issue more directly related to the present study—whether or not certain kinds of computer-based materials are superior to traditional, paper-based materials. In the case of on-line glossaries, two questions come immediately to mind: (1) would students use an on-line dictionary more than a traditional dictionary? and (2) is an on-line dictionary more effective than a traditional dictionary?

Leffa (1993) addressed these questions directly in an interesting comparative study. He claims that making a text comprehensible depends on the amount and type of support (i.e., dictionary, translation, or explanation) available for the learner. If too little support for a reading text is offered, then the text will not be comprehensible; and, he argues, if too much support is offered, then comprehension does not emerge from the text, but from the supporting material. Three criteria are essential for supported reading: (1) the support must be fast—allowing the reader to get the vocabulary item at the moment the meaning is being deduced; (2) the support must be discreet—it should not replace the text or stand between the reader and the text; and (3) the support must be contextualised—providing information that is closely related to the meaning of the word in the text (Leffa 1993: 64).

Leffa argues that the fulfilment of these criteria is nearly impossible with traditional, paper-based support tools and the computer is the logical alternative. He tested the difference in comprehension (as comprehension was operationally defined) between those who read with a traditional dictionary and those who used an on-line electronic dictionary. His results suggest that the computer-based dictionary is superior for reading tasks than a traditional dictionary because students who used the electronic dictionary took significantly less time to translate the experimental passage than did the students using a traditional dictionary. Furthermore, students who used the on-line dictionary comprehended significantly more text than did the group using a traditional dictionary.

Studies such as these suggest that the speed of the computer, if utilised, helps to increase the comprehensibility of texts for certain population. Similarly, in the present study, it is the ability of the computer to provide simultaneous reading and listening while at the same time permitting students to quickly repeat their reading and listening that should make it superior to traditional reading and listening activities or to reading-alone activities. The 'stripped hypermedia' model (my terminology) that Leffa and Hulstijn used (i.e., it was 'stripped' because it contained only one component of what hypermedia materials might make use of) will form the basis of a central argument of this thesis. The argument is this: in these initial stages of research, hypermedia should be examined in a component-by-component manner to gain an understanding of the effects of each component on specific areas of language learning. In this researcher's view this understanding should be achieved with two different types of research: (1) as with Hulstijn's investigation, there should be an attempt to understand how and why students choose to use a support tool; and (2) as with Leffa's study, a comparison should be made between traditional means of accomplishing a pedagogical goal and computer-based methods (for an overview of this kind of comparative research, see Shwalb 1994).

Understanding the components will suggest ways in which the components should be combined for greatest positive effect. This is not to say, however, that until we understand each component, classroom materials should be 'stripped'. Rather, the knowledge gained from classroom use of hypermedia materials should be integrated with results from experimentally-derived data.

2.4.2 Computer-Based Drills and Mnemonics

Coady *et al.* (1993) have also utilised the speed and customisability of the computer to teach vocabulary. In their study, students used a program that used traditional techniques of vocabulary instruction (drill and practice, selection of words from lists, recording of new words, mnemonic associations, etc.). Coady *et al.* argue that where the computer's ability to permit students to work on things in which they were interested, quiz and mark new words is utilised, significant gains in vocabulary acquisition will take place. They concluded that focused, computer-based instruction of high frequency words is beneficial for increased vocabulary size and increased reading comprehension. Again, it appears that the key factor is speed and individualisation. Coady *et al.*(1993: 222) conclude by writing that,

> The computer is an ideal instrument, due to its individualizing capabilities: students work on what they want to, and at the pace they prefer.

2.4.3 Listening and Reading with Hypermedia

To this researcher's knowledge, no other study has examined the effects of simultaneous computer-based reading and listening on second language vocabulary acquisition. There has, however, been some research with aural computer interaction.

Davidson, *et al.* trace the history of this kind of research to Atkinson (1974), where mainframe computers allowed users access by way of telephone links, asking students questions orally while students responded on a keyboard. Davidson *et al.* also cite Wise *et al.* (1989) who experimented with computerised aural feedback (when the user indicated she/he did not recognise the word the computer 'pronounced' it) and Reitsma (1989) who used touch screens to allow students to touch words on screen, and listen to an isolated word pronounced by a pre-recorded voice on the computer. All three of these studies demonstrated positive results and established, to a degree, the effectiveness of voice-assisted reading for comprehension. These studies, do not, however, address the question of vocabulary acquisition directly, nor do they make use of hypermedia equipment and software.

2.4.3.1 Experimental Questions

Several experimental questions are immediately obvious at this point: First, what are the effects of computer-based listening while reading on incidental vocabulary acquisition? Does the addition of audio information increase or decrease understanding of unfamiliar words? Does the kind of audio information (e.g., computer source or human source) affect this comprehension? Is memory of the context improved (and therefore the potential to infer from context improved) by the addition of an aural context?

Second, how does the type of computer interface affect readers? Would students rather listen to a live teacher? Are users confused by the use of the mouse;

does their relative familiarity with the computer affect their attitude towards the reading/listening experience?

Third, how do students perceive their own understanding of both new words and the text that they are asked to read? Do students who listen and read with the computer feel that they have understood more than those who read without listening? How does the language level of the learner effect performance of learners using a reading/listening program?

Finally, what additional features might be added to or subtracted from a reading/listening program to increase the positive effects of the experience? It is these four questions that comprise the focus of this thesis. However, one might also consider the effects of computer-based listening while reading on general reading/listening comprehension; on attitudes towards reading; and on long term memory of textual content and vocabulary. However, these questions while relevant, are beyond the scope of the present study.

2.5 Summary

The status of vocabulary acquisition in EFL is changing. The increased understanding of the L1 mental lexicon and the appreciation of the nature and influence of vocabulary in reading comprehension have contributed to the new alleviated position of lexis in pedagogical practices. The potential for using the computer to assist in the acquisition of vocabulary has only just begin. Studies with on-line glossaries suggest that the computer is superior to traditional methods of learning word meanings from dictionaries. Studies in incidental vocabulary acquisition suggest that a large number of words are learned over time from reading. The question remains, then, how the computer might be used to enhance incidental acquisition of vocabulary and possible decreasing the time needed to acquire word meanings from context. It is proposed that the computer's ability to provide simultaneous listening while reading could result in increased vocabulary gains over a short period of time—something not seen in reading without listening studies. The following chapter will describe the experimental method used to test this (and other) proposals.

Table of Contents: Chapter 3

CHAPTER THREE
EXPERIMENTAL DESIGN

Introduction

Chapter 1 demonstrated that the modern computer is an appropriate tool for testing the effects of listening while reading on vocabulary acquisition. In Chapter 2, we discussed the literature of vocabulary acquisition as it related to reading while listening. Chapter 3 will describe how the computer and its software was designed to test the effects of reading while listening on vocabulary acquisition.

The experiment consisted of two parts[1]. First, one group of students, divided into control and experimental groups, read (control group) and read while listening (experimental group) to a short story then took a multiple choice post-test. The test sought to measure the effects of reading while listening on vocabulary acquisition. Second, another group of students, divided into control and experimental groups, read (control group) and read while listening (experimental group) to the short story then completed a questionnaire. The questionnaire was designed to measure the effects of reading while listening on students' attitudes, ratings of comprehension and difficulty, and also to determine their perceptions of the materials. It is

---

[1] The experimental design of this chapter was described in part in a paper given by the author at EuroCALL '93.

hypothesised that the independent variable (sound) will produce significant differences between groups in post-test means and questionnaire responses.

Section 1 of this chapter will briefly review the arguments suggesting why reading while listening might produce an effect on vocabulary acquisition (in the post-test) and comprehension and attitudes (in the questionnaire). In Section 2, the experimental method will be described in detail. Section 3 will discuss the vocabulary test, the hypotheses that the test is designed to examine, and the difficulties with its design. This will be followed by a similar discussion of the questionnaire, where its design, hypotheses, and difficulties will also be delineated. The last section, Section 4, will state the statistical procedures that were used in the next chapter to analyse the data.

3.1 Preliminary Discussion

As discussed in Chapter 2, first language research suggests that there are several key factors at work in the processing, recognition, and retention of any lexical item. One of these primary factors is the acoustic nature of words. It is possible that acoustic information also provides the L2 learner with significant information and where acoustic information is provided, the learner would be better able to infer the meaning of a word (through relative stress and rhythm) and also would be more likely to remember that word (as learners would experience the written/aural context in which it was heard, and the orthographic-to-phonologic relationship of the citation form of a word).

Since many researchers have found that reading (and learning from the context it provides) is a likely source of vocabulary expansion for non-beginner adult learners, a reading and listening programme is envisioned as an enhanced context for vocabulary development. In simple reading tasks, it has been shown in the previous chapter that students can learn words after several exposures to a new or unfamiliar word—the more exposures to the new word, the more likely it is that the word would be acquired in some form.

Considering this background, there are several key questions that are being asked in this thesis. The first question is whether reading while listening would enhance the context of the materials to the extent that significant gains in vocabulary acquisition (as it will be operationally defined shortly) would be measured after just one or two exposures to a new or unfamiliar word. The second question is whether the amount of listening that students do in a student-controlled reading situation correlated to the acquisition of new or unfamiliar words as measured by the post-test (i.e., if students can control the amount of listening they do, and we can keep track of their actions, those who listen more might be expected to acquire more or fewer words than those who listen less).

The question of vocabulary acquisition is not the only concern in the present study. Students' perceptions of their own comprehension are considered a reflection of the usefulness of a reading and listening program. Similarly, from a CALL perspective, it is essential to investigate the effects of the specific program on student enjoyment, perceptions of difficulty, attitude towards using the computer for reading, and student expectations of a computer-based exercise. An investigation of student criticism of the program is considered crucial if future implementations of reading and listening programs are to be successful.

The hypermedia-equipped computer should be a satisfactory means of presenting the student with reading and near simultaneous listening while still maintaining high quality sound and readable texts. Put more simply, the computer is easier to use for reading and listening than traditional methods. For example, the student can use the computer without the pressures of peer/teacher observation and can proceed at his/her own pace. With a tape recorder, it is difficult to cue and review the tape, matching the place on the tape to the place in the tapescript. The computer, on the other hand is not constrained by linear tapes and paper tapescripts. In the research method to follow, an attempt has been made to make the navigation and listening as easy, natural, and as intuitive (for the user) as possible.

It should also be noted that studies comparing different methodologies often fail to produce significant differences in post-tests because the different methodologies are presented by a teacher (or a number of different teachers), whose actions in the classroom cannot be easily controlled by the researcher. The computer, on the other hand, can be seen as a relatively consistent source of input because the possibilities within each program are limited. The consistent character of computer assisted instruction is an asset in the research design. In the present example, the students' choices were very limited and thus the variability of the *teacher* (computer) should not be considered a liability to the experimental design.

3.2 Sample Overview

3.2.1 Population and Size

The population under examination in this thesis was post secondary school Japanese English language learners. To examine the effects of a computer-based reading activity on this population, 59 students were recruited to participate in the multiple choice component of the research study and 43 students recruited to complete questionnaires. The computer exercise and testing took place over two summers (1993 and 1994) at the Institute for Applied Language Studies at the University of Edinburgh. The sample consisted wholly of Japanese university students (ages 18-22) with the exception of one subject who was a chaperone for one of the groups of the summer of 1993. The majority of the students had come to the Institute under a group contract. However, some were recruited from the General English and Medical English programmes. Students were first approached by the researcher in their classes (with the permission of teachers and accompanying leaders). They were told (in English) that the researcher was interested in Japanese students learning English, in computers, and in reading and listening. The test was referred to as a 'quiz' in order to minimise the effects of test anxiety and also to encourage students to participate. Students were not told that the researcher was interested in vocabulary nor that the test was a multiple choice format. For the sake of incentive, and in keeping with the Japanese custom of expressing gratitude through gifts, the students were offered a small present if they participated (usually chocolate or other sweets). Students signed up voluntarily and participated in their free time.

3.2.2 English Ability and Pre-Testing

Everyone in the sample had at least six years of English instruction prior to their enrolment at the Institute. Japanese students usually come to the Institute in groups from their respective universities. In some cases, these groups are not given entrance tests or are given oral interviews. It is therefore not possible to estimate the general language proficiency of every subject. In fact, pre-test scores were available for about half ($N = 30$) of the sample who participated in the vocabulary test. These show a mixed level of English language ability—with a majority of those tested being at the lower-intermediate/upper elementary level (as classified by the students' entrance test to the teaching institute).

3.2.3 Previous Computer Experience

Unfortunately, the sample who participated in the vocabulary test were not asked about previous computer experience in any detail. After approximately the first 20 subjects completed the exercise, it became apparent that many were having difficulty with the mouse. Consequently, in order to get a broad indication of computer familiarity, the rest of the students were asked, (yes or no), if they had ever used a mouse before the experiment. Forty-three students (from a total sample size of 59) were asked. Of the 43 students asked if they had ever previously used a mouse, 60% indicated they had not (N= 26). Of those who had used one, there were some who commented that they had used a mouse "only a little". The experimental group (in both post-test and questionnaire students) were required to use the mouse much more intensively (in terms of required frequency of use and manual dexterity, pointing at relatively small areas of the screen), than the control group. It was much more important, therefore, that experimental group students use the mouse proficiently. In the experimental group, 21 students were asked if they had ever used a mouse before (72% of the total experimental group) and of these 21, 62% said they had never used a mouse before (*N*=13). Again, a *yes* answer does not necessarily mean that students felt comfortable with the mouse. These data do not show conclusively that students in the vocabulary test group were inexperienced mouse users. However, if taken with the more precise data of the questionnaire group (who came from a very similar population), we may assume that students in the experimental group were generally not experienced mouse users.

3.2.4 Randomisation

Once students had signed up for the experiment, they were randomly assigned to the control or treatment groups. Students were never asked their preference for reading or reading/listening, nor were they told until after the experiment that there was another group who were listening (or another group who were only reading). Generally speaking, the students only knew that they would be using the computer, how long it would take, and that they would be asked some questions afterwards. It is inevitable, however, that students reported back to their classmates what they were asked to do. In spite this, the students would have little motivation to discuss the contents of the test with classmates, since their test scores would not be reported to their teachers. This argument does not rule out the possibility that some of the students might have had some pre-existing knowledge of the contents of the test.

3.3 Method

3.3.1 Designing the Talking Story

The computer interface can be designed to produce sound in several ways. For example, the designer could choose to have users push buttons, click on text, use a touch screen, or just follow along as the whole text is read. There are several considerations here, some of them practical, others theoretical. The following section will examine some of these issues.

3.3.1.1 The Production of Sound

Computers can either synthesise speech by producing words by combining sounds from a bank of phonemes or, they can digitise speech through recording sound as a string of digits (Davidson, Coles, Noyes, and Terrell 1991). Hypermedia programs are more suited to digitised speech than synthesised speech. Language learning demands a high degree of natural intonation and stress and the technology of synthesised speech is not yet advanced enough to provide either.

3.3.1.2 Sound and Software

Most hypermedia programs provide the necessary programs to link sound to text (usually in the form of pre-written programming scripts). With adequate knowledge of the hypermedia program, and even without knowing the application's programming language, almost anybody would be able link sound to text simply by 'pasting' the script in the appropriate place. The process can be completely automated (save for the actual recording) if someone familiar with the programming environment pre-programs the routine to an on-screen button. The author would then simply have to use the recording program to record the sound (which is as easy as using a tape recorder), choose the text/graphic to be linked and then type in the name of the sound file. These kinds of pre-programmed routine are readily available in the public domain for most hypermedia programs.

3.3.1.3 Linking Sound and Text

Sound is linked through text with specialised software. An extensive review of appropriate software is not within the scope of this thesis (see Appendix 1.1 for a brief overview), but there are several considerations that require discussion. First, software development is market driven and the educational market is relatively small compared to the business market. Since software for development of hypermedia projects is outside the mainstream of business uses, it is slightly behind other kinds of software such as word-processing, spreadsheets and databases.

One can embed sound in text with almost any *Windows* software. The embedded sound file is usually indicated by a small icon in the body of the text indicating that the user can listen if s/he so desires. Specialised software such as *Guide* permits the embedding of sound without the icon. The sound can be attached directly to a word or it can be attached to a graphical button. It is not known how to inform the user efficiently of various computer-based features. If the text is full of icons, one might argue that the reading would be 'unnatural' because there are so many visual interruptions. If the sound is attached to a word and the word is not highlighted in some way, then students would have uninterrupted text but might not be aware that they could also listen. In any case, clear instructions and time to experiment are crucial elements of this experimental method.

3.3.1.4 The L1 Reading Experience

The goal of this study is, as much as possible, to mimic the positive experiences that many good L1 readers enjoyed in their childhood of reading quietly along with a parent, i.e., students read in comfort, they read at a leisurely pace and without the pressure to perform that might occur in a teacher-student reading situation. However, we do recognise that reading with a computer is not the same as reading with a parent at bedtime. There are many differences and many psychological, emotional, and developmental factors that make the L1 reading experience with a parent fundamentally different from L2 reading while listening with a computer. Nevertheless, teachers of first language reading, who use computers and talking books, go to great lengths to make the experience as book-like and as tactile as possible. In the normal (non-experimental) teaching situation, the best method for allowing navigation through a text would be a touch screen (cf. Chomsky 1990). Screens in classrooms sometimes display photographs of an open book and may even include the 'reader's' hands turning the pages. The teaching situation also allows for various other computer-based activities such as choices between plain text, text that highlights itself as the spoken text is read, and subtitles that follow the voice. Stated goals in studies like Chomsky's (1990:39) are that computer-based reading materials should "...capture the quality and charm of the real-life reading experience." Goals such as these emphasise the importance of positive developmental characteristics for children of reading while listening with an adult. While the developmental and educational needs of adults learning a second language are different from a child learning to read in his/her first language, there can be little doubt that in a non-experimental context, charm, quality, and simplicity are highly desirable for L2 learning as well.

3.3.1.5 Experimental Reliability

Many of the features that make a program valid in the classroom have too many variables for the experimental context where reliability is of primary concern. Furthermore, hardware and cost constrain the possibilities within most experimental contexts (e.g., motion video, laserdiscs, touch screens, etc. are still prohibitively expensive for large scale use). However, it is hoped that the experimental method discussed here was a compromise between a sterile experimental procedure and a program that could be used without alteration in the classroom.

For this reason, the reading was intended to appear familiar and intuitive to students (by being book-like) while, for experimental purposes, controlling as many variables as possible. The following table summarises the key characteristics of such an exercise:

Table 3.3-1

Design Considerations in an Experimental and Classroom Context

experimental considerations		classroom considerations	
description	rationale	description	rationale
no hypertext and multimedia limited to sound only	effects can be attributed to the variable of sound	the text flows from page to page and permits unlimited linear navigation.	mimics a book in its accessibility
the text is unbroken by on-screen buttons (apart from the two navigation buttons)	no advantage given to experimental group	the voice is recorded and not synthesised	sound is natural and not machine-like
no other program features such as word practice, highlighted text, or subtitles	effects can be attributed to the variable of sound	the interface is easy and requires minimal training	students not frustrated
the interface is easy and requires minimal training.	previous computer experience should not be a variable	users in the experimental group can choose to use the listening facility or to read without it.	students have a degree of freedom to listen or not. The offer of choice is seen to be more representative of a typical self-access or classroom situation.
users control the sound with the mouse	mouse considered to be a familiar means of interacting with a computer screen	students should be able to listen to as many individual words (in citation form) as possible	accommodate the curiosity of students and provide them with further opportunity to clarify words

It was thus decided that the spoken sentence would be elicited from the computer by *clicking* on (placing the on-screen cursor with the mouse on an appropriate point, then pushing the mouse button down) the first letter of every sentence. Similarly, individual words can be heard by clicking on any part of a word with a sentence. Students were told that if the cursor changed to a right-facing arrow as it was 'dragged' over a word, then the word could be clicked-on and heard.

While students would require some training to be able to do this, the resulting screen would be simple, visually unbroken, and least as familiar as the screen of, for example, a word-processor. It was assumed that most students would be at least superficially acquainted with the mouse but that even those without computer experience would become proficient in its manipulation.

3.3.2 Technical Requirements

Sound puts considerable strain on computer resources. Hult *et al.* (1990) estimate that one second of digitised sound takes 7-11 kilobytes of hard disk space. Ten minutes of sound on this system would therefore consume 6-7 megabytes of space (or roughly the size of six high density diskettes). The present research shows that considerably more disk space than that is required on our systems. Using the very highest setting on PC hardware (that produces stereo sound at CD quality), 1000 words of text takes about thirty-five megabytes of disk space for the sentences and about seven megabytes for the citation-form words. It is estimated that the story has about 15 minutes of recorded speech.

Considerably less space can be taken if the sound is not recorded in stereo, at the highest setting. Larger hypermedia documents could expand rapidly in size if sound is used too freely throughout the materials. Furthermore, long segments of sound require several megabytes of RAM. On a PC running *Windows*—which in itself requires a relatively recent computer, there are fewer problems with RAM due to system features that were not available to Hult *et al.* at the time. Furthermore, as an indication of how quickly things change, when the present research began, the computers could just manage that size requirements of the experimental text. By the time the data collection was finished, however, we were able to use computers where the sound files only took up a fraction of the overall disk space.

3.3.3 Text/Navigation

The text was a short story consisting of about 1000 words called *The Blue Bouquet* (see Appendix 2.1). There was some degree of difficulty in finding an appropriate story for different levels of student with possibly varying interest. *The Blue Bouquet* was chosen intuitively by the researcher and a colleague (Dwyer 1992) long experienced with Japanese students. The following section reports some of the decision criteria used.

3.3.4 Comparisons with Day **et al.** *(1991)*

As we said in the previous chapter, the present research was partly modelled on Day *et al.* Some key elements were taken from that study as exemplary for the present one. These included:

- Text length
  (approximately 1000 words—found to take about 30 minutes to read in a 13 student pilot study);
- Multiple choice format for the post-test
  seen as a logical first step for a previously un-researched area and a relatively unproblematic process of statistical analysis;
- Text chosen for its interest to students
  the text was chosen by the researcher and a colleague who also teaches Japanese students short-listed 3 texts and then chose the 'most interesting';
- Text chosen at the appropriate level for first or second year Japanese university students
  the level was decided by the researcher and the same colleague mentioned above. The chosen text was then used in the pilot study where students were subsequently queried about its appropriateness;
- Control and experimental group design
  this design was thought to be a logical first step in a study that had few precedents.

There were several changes to the basic design of Day *et al.*. These changes were made either as improvements or for practical reasons. Among these were:

- An authentic, un-altered text
  it was felt that an unaltered text (both in length, and frequency of target words in context) would make the results more generalisable for a classroom context;
- No pre-test
  the number of Japanese students attending courses at the place of research was limited. This ruled out the possibility of using a group of students for a pre-test (to determine unknown or unfamiliar words) and another group for the actual experiment. It was decided that using the same students for pre-test and experiment was unsatisfactory as the time period where students were available did not allow for the pre-test to precede the experiment by sufficient time. Unknown or unfamiliar words were chosen on the basis of the researcher's and other's (Dwyer 1992) experience and intuition. It was thought that since the test was a comparison between control and experimental groups this would be an acceptable compromise;
- Multiple choice post-test
  On consultation with several experts in test composition (e.g., Davies 1992), it was decided that there need not be an 'I don't know' option on the multiple choice test. It was felt that students might take this option before they risked 'being wrong'. Similarly, no correction for guessing was made. It was felt that if students got the answer right, it was inappropriate to 'second guess' their reasons for doing so.

There were other considerations when deciding the size of the text. Firstly, text length was constrained by the amount of free disk space on the computers that were available when the experiment began. After some trial runs, we determined that 1000 words would consume about forty two megabytes of disk space. As has been said, during initial stages of the study, forty-two megabytes represented more than half the disk space of even our largest computer. In addition, the researcher was determined not to tax the goodwill of the students involved.

Black typed text appeared on screen page by page (screen by screen) on a white background. The image on the screen consisted of linearly-linked frames which the student navigated by a 'next page' button (goes to the next page in the book), and a 'previous page' button (goes to the previous page in the book). Students turn pages by clicking on these buttons with the mouse. They can skip pages without being forced to spend a fixed period of time reading a page.

3.3.5 Tools

The computer equipment consisted of various IBM compatible machines with colour monitors and a hard disk drive. The software used for the reading was *Guide 3.1* (OWL International Ltd 1992). *Guide* runs under *Windows 3.1*. Computers used for sound were equipped with high resolution sound cards (*SoundBlaster 16* from Creative Research Inc. 1992). Students listened through stereo headphones. All the computers had a mouse, mouse mat, and keyboard, although the software was designed in such a way that the keyboard was not used.

The individual computers contained various processors but there was no noticeable difference in speed with the software used. All the machines used at least 386SX processors or better (16 Megahertz with four megabytes RAM). The computers equipped with sound were 486 (33 and 50 Megahertz, with four or eight megabytes of RAM respectively).

3.3.6 Sound

Any *word* suspected to be unknown or unfamiliar (cf. section 2.1.3) was linked to a recorded version of the word in citation form. This included all the words in the post-test and several others. Every recorded *sentence* was linked to the first letter of its written counterpart. Students were able to listen to individual words or individual sentences by clicking on the word to hear the word, or by clicking on the first letter of the sentence to hear the sentence. It was possible to listen to each sentence consecutively but students were not forced to do so. Students could listen to each word or each sentence as many times as they liked.

3.3.7 Student Logs

A logging system was programmed into the exercise to provide information on which word/sentence was listened to and when pages were turned. All of the logging took place in the background. Students were unaware that the recording was taking place (see Appendix 2.2 for a sample log file).

The purpose of the log system was to determine precisely how much the listening facility was used by members of the experimental group. Since it was possible for students in the experimental group to read without listening, placement in the experimental group was no guarantee that students were affected by the independent variable (listening). The log system was thus necessary to determine how much the facility was used. It was possible, therefore, to correlate the amount of use to user post-test scores and to questionnaire results.

3.3.8 Control and Experimental groups

The post-test and questionnaires were administered to different groups. In each group, half the students were randomly placed in a control group who read without listening. The other half formed the experimental group who read while listening. Both groups were given time to learn how to turn pages and (as in the case of the experimental group) to work the sound facility. The familiarisation procedure took place on a 'dummy' text and took between 5 and 10 minutes (students were given up to 10 minutes to say when they were ready). The familiarisation period was immediately followed by the real text and the test immediately followed the reading. The text, setup, and initiation procedure was exactly the same for the post-test group and the questionnaire group.

3.4 Measurement Tools

3.4.1 Vocabulary Test

3.4.1.1 Design

A vocabulary test consisting of fifty multiple choice questions was constructed and piloted with a group of 13 Japanese students. Two test items were rejected after the pilot, leaving 48 for the experiment. For the most part, the correct answer among the distractors was adapted from *The Oxford Learner's Dictionary* (see Appendix 2.3). It was felt that students at the level represented by the population under examination (Japanese university students) would likely be able to understand definitions from this kind of dictionary. Similarly, distractors were worded (intuitively) so as to not be to difficult in themselves. Since the present research is concerned with the learning of words from context, and many of the words were only encountered once, it was decided that the correct answer would be the definition of the word as it appeared in experimental context, even if that definition deviated from the word's prototypical meaning (i.e., the first meaning of the word given in the dictionary).

As far as possible, the test adhered to established multiple choice testing practices. Several rules were seen as particularly critical (adapted from Lado 1961):

1. The alternative responses for each item should be kept as brief and as equal in length as possible.
2. The items should be kept as independent from each other as possible (answers to one should not unwittingly provide the answer to another item).
3. The distractors and correct answers should be as easy to understand as the test item (cf. section 3.2.4 above).

Three distractors and the correct answer accompanied each item. The placement of correct answers was randomised and instructions to the test were given in both Japanese and English at the head of each page. Students in both groups were also orally reminded (in English) to think of the story when considering the correct answer for each item.

3.4.1.2 Vocabulary Test Hypotheses

For the purposes of Hypotheses 1 and 2, 'vocabulary acquisition' was operationally defined as the student's ability to choose the correct definition of the word from a list of alternatives. The vocabulary post-test was designed to test the following null hypotheses:

3.4.1.2.1 Null Hypotheses 1 and 2

1. There will be no significant difference between the control group and the experimental group in the mean number of target words answered correctly on the post-test.
2. There will be no relationship between the amount of listening a student does and his/her post-test score.

3.4.1.2.2 Hypotheses 1 and 2

From these null hypotheses, the following hypotheses are offered:

1. The difference between the mean post-test scores of the control group and experimental group will be significant.
2. There will be a relationship between the amount of listening a student does and his/her post-test score.

3.4.1.3 Limitations to this Design

3.4.1.3.1 Multiple Choice Tests for Measuring Vocabulary Acquisition

In the past, testing vocabulary acquisition after learners' exposure to several different methods has proved controversial on two levels. The first is that measurements used to determine the differences in vocabulary acquisition are rarely neutral between methods. The second is that these types of test only measure the state of vocabulary at a given point in time or over a very short period (Meara 1989). Meara argues that matrix models that incorporate the mathematical probability of retention or loss of a word might be used to predict that changes in a student's vocabulary. Furthermore, Meara advocates a test where students decide which words they know and which words they do not know, arguing that students are "usually fairly conservative about whether they know words or not, and they tend to underestimate their abilities rather than to over-estimate them." (p. 72). This kind of test is seen as a more objective, neutral measurement when disparate methods are being compared.

The multiple choice test used in this experiment can also be defended on two levels. First, both methods under examination in this study are essentially 'learning from context' methods. When both groups are essentially an examination of incidental acquisition we can reasonably assume that there is no advantage given to one group or the other by the type of testing. Second, the multiple choice format was chosen because it has been used extensively on previous similar studies (e.g., Day *et al.*1991) and because its design, use, and subsequent analysis are fairly well understood and reviewed in the literature.

In the present study, we accept that the type of learning that a one-off multiple choice test measures does not reflect the overall nature of vocabulary retention and loss. However, without more substantive support in the literature for more sophisticated measurements (like Meara's matrix model), the one-off, multiple choice format seems a logical starting point for the testing of null hypotheses 1 and 2. However, in a future study, after the data from this study are analysed and discussed, it would also be logical to repeat the investigation with different kinds of testing and using different models—models that more accurately reflect the dynamic nature of the second language mental lexicon. That kind of study, while interesting and no doubt essential for a fuller understanding of vocabulary acquisition, does not fall within the parameters of the present study.

3.4.1.3.2 Receptive and Productive Knowledge of Words

The multiple choice test does not require students to use words productively nor does it require them to demonstrate their knowledge of its association to other words. It was thought that it would be too much to expect students to acquire productive skills with new or unfamiliar words after encountering them only once or twice in context—enhanced or not.

3.4.2 Questionnaire

3.4.2.1 Questions of Experience

The questions pertaining to students' previous computer experience deserve some explanation. There are several ways the question of previous experience could be phrased. For example, the students could be asked to rate their confidence with the computer and the mouse, the students could be asked if they used computers at home or at university, etc.. The present format was chosen for two reasons. First, the researcher's own experience of teaching novice users how to use computers revealed that frequency of use was the primary factor in determining a user's ability to interact with a computer. If students used a computer regularly, they would be familiar with different computer skills such as pointing and clicking, screen quality, and the kind of programs that they might expect to encounter. Second, the researcher was aware that computers are used differently and with different operating systems throughout the world. Computer experience does not guarantee mouse experience, nor does mouse experience guarantee many kinds of computer experience. It was felt that the question should be phrased to accommodate this distinction and that frequency of use over given time periods would give the best indication of the user's interface skills.

3.4.2.2 Questions of Plot and Content Difficulty, Vocabulary, and Computer Interaction

In the course of investigating the performance of the two groups on the vocabulary test, and discussing students' perception of the exercise, it was noticed that students who listened to the short story almost always commented on the pleasure of the experience and the fact that it was easy to understand because they could listen at the same time. It was decided that student perception of the program should be investigated in the questionnaire. The purpose of these items was to examine the perceptions of difficulty that both groups had, i.e., did students who read and listen think 'more positively' about the level of difficulty of the story, the words, and the computer than those who read the story without listening? Each item on the questionnaire was designed to answer a different aspect of this overall question. It was decided for consistency and to avoid confusion, all the questions would be phrased positively. An attempt was made to keep the questions neutral and not leading, although some questions admittedly were aimed at examining extreme reactions to the experiment (e.g., 'waste of time' item).

It is possible that students who had to use the mouse more (i.e., the experimental group) would experience higher frustration than those who only need the mouse for turning pages, especially if students were unfamiliar with mouse use. Generally speaking, frustration with the computer itself is overcome with experience (Stevens 1991 and cf. section 1.4.3.1). The degree to which previous experience affected the enjoyment and usefulness of the experience is one focus of the questionnaire.

3.4.3 Questionnaire Hypotheses

3.4.3.1 Hypotheses Relating to Between Groups Differences

We proposed that the listening would have several effects on the experimental group including an effect on perceived difficulty of the plot and vocabulary, enjoyment, difficulty of mouse manipulation, and preferences for reading materials. The questionnaire tested the following null hypotheses; the alternative hypotheses is stated below each null hypothesis.

3.4.3.1.1 Null Hypothesis 3a

> There will be no significant difference between control and experimental group ratings of vocabulary difficulty.

3.4.3.1.2 Hypothesis 3a

There will be a significant difference between control and experimental group ratings of vocabulary difficulty.

3.4.3.1.3 Null Hypothesis 3b

There will be no significant difference between control and experimental group ratings of plot difficulty.

3.4.3.1.4 Hypothesis 3b

There will be a significant difference between control and experimental group ratings of plot difficulty.

3.4.3.1.5 Null Hypothesis 4

There will be no significant difference between control and experimental group ratings of program enjoyment.

3.4.3.1.6 Hypothesis 4

There will be a significant difference between control and experimental group ratings of program enjoyment.

3.4.3.1.7 Null Hypothesis 5a

There will be no significant difference between control and experimental group ratings of plot comprehension.

3.4.3.1.8 Hypothesis 5a

There will be a significant difference between control and experimental group ratings of plot comprehension.

3.4.3.1.9 Null Hypothesis 5b

There will be no significant difference between control and experimental group ratings of vocabulary comprehension.

3.4.3.1.10 Hypothesis 5b

There will be a significant difference between control and experimental group ratings of vocabulary comprehension.

3.4.3.1.11 Null Hypothesis 6

There will be no significant difference between control and treatment group ratings of mouse difficulty.

3.4.3.1.12 Hypothesis 6

There will be a significant difference between control and treatment group ratings of mouse difficulty.

3.4.3.1.13 Null Hypothesis 7

There will be no significant difference between control and experimental group preference for the reading medium.

3.4.3.1.14 Hypothesis 7

There will be a significant difference between control and experimental group preferences for the reading medium.

3.4.3.1.15 Null Hypothesis 8

There will be no significant difference between control and experimental group desire to use similar materials again.

3.4.3.1.16 Hypothesis 8

There will be a significant difference between control and experimental group desire to use similar materials again.

3.4.3.1.17 Null Hypothesis 9

There will be no significant difference between control and experimental group perception of the computer as a language learning tool.

3.4.3.1.18 Hypothesis 9

> There will be a significant difference between control and experimental group perception of the computer as a language learning tool.

3.4.3.2 Hypotheses Relating to Subjects' Predispositions

It is possible that students began the reading exercise with a dislike for computers, or a dislike for reading in English or in Japanese. They may find reading in English (with any medium) difficult and frustrating. If any or all of these are the case, it is essential to know if there is any relationship between students' predisposition and their opinion of the experimental exercise. The questionnaire included several items designed to determine the relationship between students' general attitudes and their attitudes to the immediate task. To determine the relationship between these predispositions and students opinion of the experiment a non parametric correlation was performed between the relevant questionnaire items. The primary purpose of this correlation was to rule out the possibility that students, on the whole, answered according to their predisposition rather than their immediate opinion. With this in mind, we assumed that the control and experimental groups were equally likely to be affected by their predispositions and the correlations were performed on the combined group—not the control and experimental groups separately. Of course it is also possible that the experimental group were more or less affected by their predispositions than the control group and that the correlations between items were different in each group (students who do not like reading, for example, might have liked the computer reading if they were in the experimental group). However, the questionnaire was not designed to examine the differences between groups in this way. Two hypotheses (Hypotheses 11a and 11b), however, dealt specifically with predispositions that might affect the experimental group only. The following seven null hypotheses were proposed to examine questions of students' predisposition:

3.4.3.2.1 Null Hypothesis 10

> There will be no significant correlation between how much students like computers and how much they liked the computer-based reading exercise.

3.4.3.2.2 Hypothesis 10

> There will be a significant correlation between how much students like computers and how much they liked the computer-based reading exercise.

3.4.3.2.3 Null Hypothesis 11a

> There will be no significant correlation in the experimental group between students' enjoyment of listening while reading in general and their enjoyment of the computer task.

3.4.3.2.4 Hypothesis 11a

> There will be a significant correlation in the experimental group between students' enjoyment of listening while reading in general and their enjoyment of the computer task.

3.4.3.2.5 Null Hypothesis 11b

> There will be no significant correlation in the experimental group between students' enjoyment of listening while reading to a teacher and their enjoyment of the computer task.

3.4.3.2.6 Hypothesis 11b

There will be a significant correlation in the experimental group between students' enjoyment of listening while reading to a teacher and their enjoyment of the computer task.

3.4.3.2.7 Null Hypothesis 12

There will be no significant correlation between students' enjoyment of short stories in Japanese and their enjoyment of the computer-based reading.

3.4.3.2.8 Hypothesis 12

There will be a significant correlation between students' enjoyment of short stories in Japanese and their enjoyment of the computer-based reading.

3.4.3.2.9 Null Hypothesis 13

There will be no significant correlation between students' enjoyment of short stories in English and their enjoyment of the computer-based reading.

3.4.3.2.10 Hypothesis 13

There will be a significant correlation between students' enjoyment of short stories in English and their enjoyment of the computer-based reading.

3.4.3.2.11 Null Hypothesis 14

There will be no significant correlation between students'' difficulty with reading in English in general and their difficulty with the computer-based reading.

3.4.3.2.12 Hypothesis 14

There will be a significant correlation between students'' difficulty with reading in English in general and their difficulty with the computer-based reading.

3.4.3.3 Hypothesis Relating to Usefulness of the Program

To determine if the students felt that they overall experimental reading was useful or not, we posited the following null hypothesis:

3.4.3.3.1 Null Hypothesis 15

There will be no significantly different response between groups that the computer-based reading is useful.

3.4.3.3.2 Hypothesis 15

There will be a significantly different response between groups that the computer-based reading is useful.

3.4.3.4 Hypotheses Relating to Previous Experience

To determine whether previous experience was related to enjoyment, interface difficulty, and comprehension we offered the following three null hypotheses:

3.4.3.4.1 Null Hypothesis 16

There will be no significant correlation between previous computer experience and enjoyment.

3.4.3.4.2 Hypothesis 16

There will be a significant correlation between previous computer experience and enjoyment.

3.4.3.4.3 Null Hypothesis 17

There will be no significant correlation between previous computer experience and interface difficulty.

3.4.3.4.4 Hypothesis 17

There will be a significant correlation between previous computer experience and interface difficulty.

3.4.3.5 Questionnaire Design Limitations

The questionnaire consists of two parts. The first is a series of closed questions that require students to agree or disagree on a five-point Likert scale with a series of positive questions. It was felt that a questionnaire would be less intimidating than the post-test and would therefore provide the researcher with a measurement tool that would counterbalance the somewhat strenuous and pressured experience of those who took the multiple choice test. Thus, the questionnaire does not require the students to recall the meaning of the text, specific vocabulary, or syntactic structures. A closed (as opposed to an open-answered) section with degrees of agreement/disagreement was included on the grounds that it would allow for a systematic tabulation of results, consistent interpretation of the questions by students, and because this type of questionnaire has been found to be less frustrating for respondents than other types or survey (Mouly 1978).

The second part of the questionnaire consists of several open-ended questions designed to evaluate the program in terms of usability in a teaching context and potential improvements for such a use. This part of the questionnaire is intended to answer questions raised in Chapter 1 concerning students' expectations of computer-based reading materials, their opinions as to how the materials might be improved, and their overall impression of the activity.

The validity of any questionnaire depends on the willingness and ability of the participants to provide accurate information (Mouly 1978). With this in mind, the questionnaire was translated into Japanese—in order to minimise confusion, and to reduce the effects of higher or lower language skills on the result. The questionnaire was also piloted on several Japanese teachers and students. As with any questionnaire that requires answers from a culture different from the researcher, apparently innocuous questions may sometimes be taken as provocative or threatening by the subjects (Sudman and Bradburn 1982). For example, one of the questionnaire translators (Inoue 1992) pointed out that if students are asked to agree or disagree that reading aloud with the teacher is enjoyable they may inevitably choose to disagree. The rationale was that every Japanese student experiences reading and listening in some form during their school years—a practice that is usually disparaged and used by critics of the system as an example of ineffective Japanese English teaching methodology (the item remained in the questionnaire as its purpose was determine if any relationship between students' predispositions and their reactions to the experiment existed).

There may well be other difficulties with threatening questions of which we are not aware. The advantage of the control/experimental group design, however, is that threatening questions exist equally for both groups who share a large degree of cultural and educational homogeneity.

The questionnaire is quite short. It was felt that a short questionnaire would be advantageous as the research relied on the goodwill of the participants. If one group of students told others in their class that the questionnaire was short, painless, and in Japanese, it was felt that this would result in a larger sample. The drawback of the short questionnaire is that general attitudes are more reliably measured with a variety of related questions—one group of questions for each attitude to be measured. However, it was decided that if the questions were straightforward, and the data was interpreted carefully, the advantages of the short questionnaire outweighed the benefits of more complex designs.

Comments in the open-ended section of the questionnaire provided a possibility for evaluation of the exercise itself. It was aimed at providing insight into program improvements and students' self observations of how they used the materials.

3.5 Statistical Procedures

3.5.1 Multiple Choice Post- Test Statistics

The use of a t-test follows three assumptions regarding the data obtained from the multiple choice test (Hatch and Lazaraton 1991)

1. The data represented by the dependent variable (vocabulary acquisition) will be measured as an interval score.
2. The scores will be normally distributed. Thus the mean and standard deviation are appropriate measures of central tendency and variability.

Assuming that the data fell within these criteria, an independent samples t-test was used to test the null hypothesis that there is no difference between control and experimental group post-test scores.

A Pearson correlation was used to test the null hypothesis that there is no significant relationship between the amount of sound-facility use (which is provided from the log system) and scoring on the post-test.

3.5.2 Questionnaire Statistics

Distribution of responses of the data provided by the questionnaire was measured using a Chi-Square Distribution. Correlations between items was carried out using a Spearman procedure for measuring non-parametric correlations.

3.6 Summary

The experimental method described in this chapter consisted of two parts. The first part was a control group/experimental group design where the two groups are given identical reading exercises with the exception that the experimental group has the option of listening to the story as they read it. Following the reading exercise, students were asked to take a multiple choice vocabulary test, the results of which were examined and the null hypothesis that there is no significant difference between groups was tested with an independent samples t-test. Included in this part of the design is a description of the log system used for testing the null hypothesis that there is no relationship between the amount of listening a student does and his or her post-test score.

The second part of the experimental procedure involves a different control and experimental group reading the same materials. In this case, following the reading, the students are asked to complete a questionnaire. The questionnaire consists of 18 items with which students agree or disagree using a five-point Likert scale. The questionnaire also includes three open-ended items where students can suggest improvements, make comments about difficulties, or respond in any other way they choose. A Chi-Square Distribution was used in the next chapter to test several null hypotheses stating that the distribution of responses between categories do not differ significantly between control and experimental groups. Additionally, non-parametric correlations was used to test the null hypotheses that there was no relationship between several kinds of student predispositions and their impression of the computer-based reading. Chapter 4 will examine the statistical results from the tests in both parts of the experimental design.

Table of Contents: Chapter 4

CHAPTER 4
EXPERIMENTAL RESULTS

Introduction

Chapter 4 will describe the statistical results of the experimental procedure outlined in Chapter 3 and supported by Chapters 1 and 2. Section 1 of this chapter gives an overview of the sampling and recruitment procedures, and the subjects' language level. It also discusses the computer packages used to analyse the data. In sections 2 and 3, each hypothesis will be examined in turn. Section 2 will analyse the data from the vocabulary test. The analysis will consist of four parts: (1) tabulated summary statistics of the means, maximum and minimum scores, standard deviations, and standard error of means of the control and treatment groups; (2) graphs of the frequency distribution of scores in both groups; (3) the results of an independent samples t-test; and (4) an examination of the data from the computer-based log record for the log data will include tables of the total number of clicks by users in the experimental group, the total number of clicks on sentences, the total number on words, and the proportion of sentences and words that were used out of the total possible. This data will then be correlated to the vocabulary test scores using a Pearson Product Moment correlation.

Section 3 will examine the data from the questionnaire. The sample size and characteristics will be discussed in the same way as the previous sections. Each hypothesis relating to the questionnaire will then be examined in turn. Sections 3.3 through 3.9 will examine predicted differences between groups and section 3.10 will cover students' predicted predispositions. Finally, 3.11 and 3.12 will discuss computer usefulness, and correlations between previous computer experience and enjoyment, respectively. The last section, section 4, will discuss several unpredicted results from the questionnaire data. These will include a second look at the combined group's responses to the computer as a reading medium, their generally low level of previous computer experience, the combined group's impressions of the computer's usefulness, and the combined group's assessment of story difficulty.

Correlations will be made using the Spearman correlation to examine relationships in each group between various questionnaire items. To test hypotheses that question the independence between row and column variables (e.g., the independence of sound/no sound and enjoyment/non enjoyment) a Pearson Chi-Square Distribution will be used (see Appendix 3.1 for a mathematical description of this and other tests used in this chapter). The last section (section 5) will include statistical results that were not specifically addressed by hypotheses in Chapter 3.

Throughout this chapter, a confidence level (α) of .05 is assumed.

4.1 Computer Programs for Statistical Analysis

After marking tests and questionnaires, the data were entered into a database (*Microsoft Access*) to facilitate sorting and querying. The statistical test and analyses were carried out using the *Statistical Package for the Social Sciences* (*SPSS for Windows* Release 6.0 1993). Graphs and other figures in the following sections were produced by *SPSS* and *Microsoft Graph*.

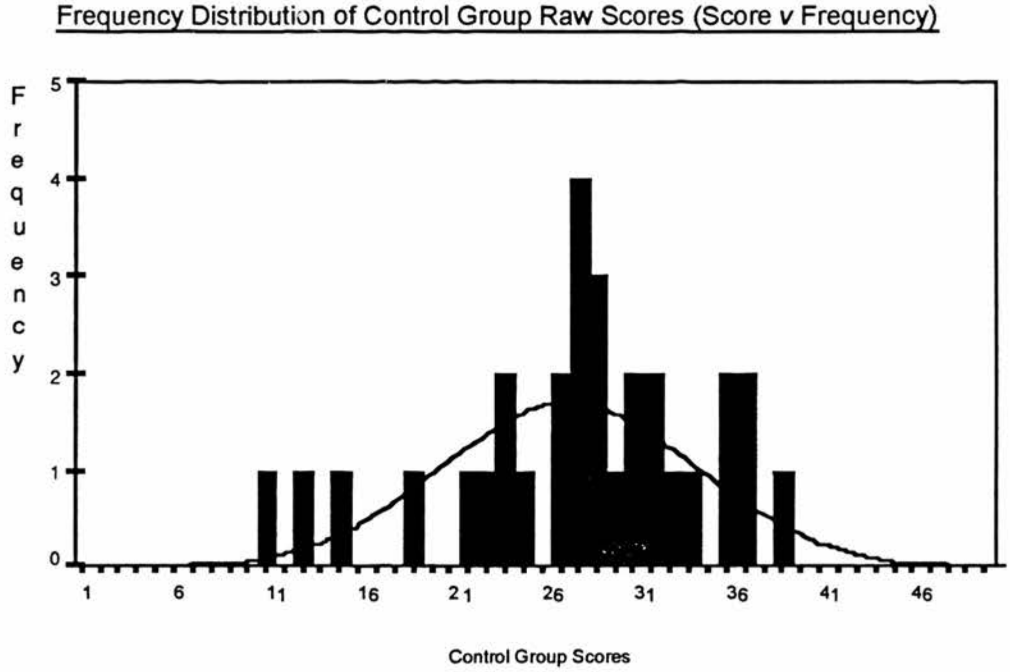
4.2 Vocabulary Test Results

The first part of the vocabulary test section will establish that both control and treatment group scores are normally distributed. Following that, each hypothesis as set out in Chapter 3 will be re-stated and the statistical procedures used to test that hypothesis will be described.

4.2.1 Control Group Distribution

There were 30 students in the control group. The number of students who achieved each score is shown in the following graph (for individual scores see Appendix 3.2):

Figure 4.2-1



The line on this graph is a superimposition of the theoretical normal distribution. As the graph demonstrates, there are no outliers in the control group sample and the scores group around the mean. Table 4.2-1 summarises some essential statistics:

Table 4.2-1

Statistical Summary of the Control Group

Mean	Std Dev	Minimum	Maximum	N
26.90	6.94	10.00	38.00	30

4.2.2 Experimental Group Distribution

The experimental group consisted of 29 individuals and also showed no obvious outliers in the scoring. Figure 4.2-2 shows the frequency of each score for the experimental group. The theoretical normal curve is displayed on the graph.

Figure 4.2-2

Frequency Distribution of Experimental Group (Score v Frequency)

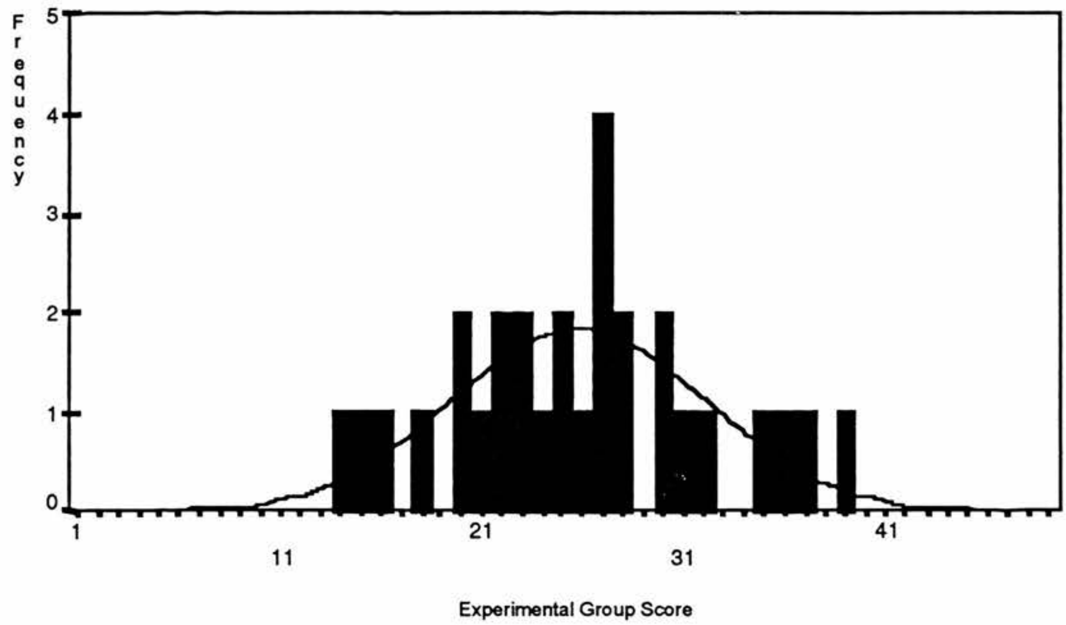


Table 4.2-2 summarises some essential statistics of the experimental group:

Table 4.2-2

Statistical Summary of the Experimental Group

Mean	Std Dev	Minimum	Maximum	N
25.79	6.43	14.00	39.00	29

As these tables show, the standard deviations were similar for both groups. In section 4.2.3, we will discuss the differences in means between both groups.

4.2.3 Hypothesis 1 and 2

Hypotheses 1 and 2 from Chapter 3 dealt with the difference in post-test scores between the control and experimental groups. Hypothesis 1 from Chapter 3 states that:

The difference between the mean post-test scores of the control group and experimental group will be significant.

4.2.3.1 Results of the t-test

To test the null hypothesis, an independent samples t-test was performed using the above data. The next table summarises the results from this test:

Table 4.2-3

Results of Independent Samples t-test

Group	*N*	Mean	Mean Difference	SD	SE of Mean	df	t-value	2-tailed sig.
Control	30			6.94	1.27			
			1.11			57	.64	.53
Experimental	29	25.79		6.43	1.19			

As Table 4.2-3 shows, there was a very slight difference between group means and the t-test demonstrated that it is not possible to reject Null Hypothesis 1.

4.2.3.2 Use of the Listening Facility

The computer-based logs record which sentence or word was clicked on and at what time. The data from these records were obtained by adding up the number of clicks, and counting the repeated sentences or words.

4.2.3.2.1 Data Summary

The log data from 3 experimental group subjects were lost due to technical difficulties on the day of their testing. Nevertheless, it was possible to count the amount and kind of use of 26 subjects and to correlate this use to the subjects' test scores (see Appendix 3.3 for raw data of different kinds of clicking). Four kinds of user action will be examined: (1) the total amount of mouse use (not including page turns) of each student; (2) the proportion of individual sentences to which the students listened (out of a possible 97); (3) the proportion of individual words to which students listened (out of a possible 50); (3) the number of repeated sentences; and (4) the number of repeated words. Table 4.2-4 summarises the overall use of the listening facility by the experimental group, including the mean number of times the mouse was clicked, the proportion (where applicable) of the possible number of clicks, and the minimum and maximum number of clicks and the number of students in the sample.

Table 4.2-4

Summary of Use: Listening in the Experimental Group.

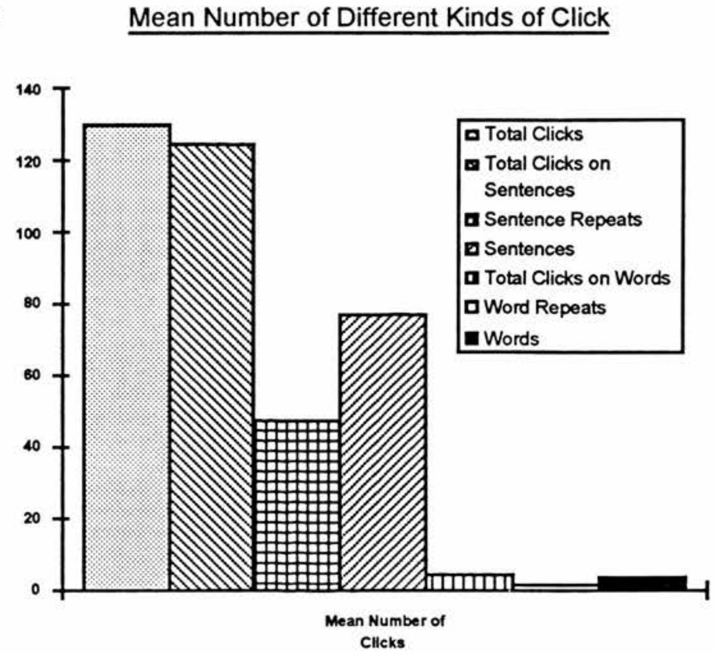
Variable	Mean	Proportion	Min	Max	N
Total Clicks	129.46	N/A	22.00	296.00	26
Total Clicks on Sentences	124.62		7.00	292.00	26
Sentence Repeats	47.50	N/A	.00	195.00	26
Sentences	77.12	79.05%	6.00	97.00	26
Total Clicks on Words	4.85	N/A	.00	30.00	26
Word Repeats	1.35	N/A	.00	13.00	26
Words	3.50	7.0%	.00	17.00	26

(1 click = 1 use of the mouse)
(N/A = Not Applicable)

Table 4.2-4 shows the total number of clicks on words and sentences and out of that total, the number that were repeats. In the *total clicks* category, the number given is the average number of clicks performed by students in the experimental group. This number includes clicks on words, clicks on sentences, and repeated clicks. It is therefore a gross account of the amount of listening done by students. *Total clicks on sentences* is the same gross account with the number of word clicks and repeats subtracted from the *total clicks* category. To get the number of sentences clicked on out of the total possible, *sentence repeats* was subtracted from *total clicks on* *sentences.* The same procedure was followed for *total clicks on words* and *words.*

The following graph illustrates the differences:

Figure 4.2-3



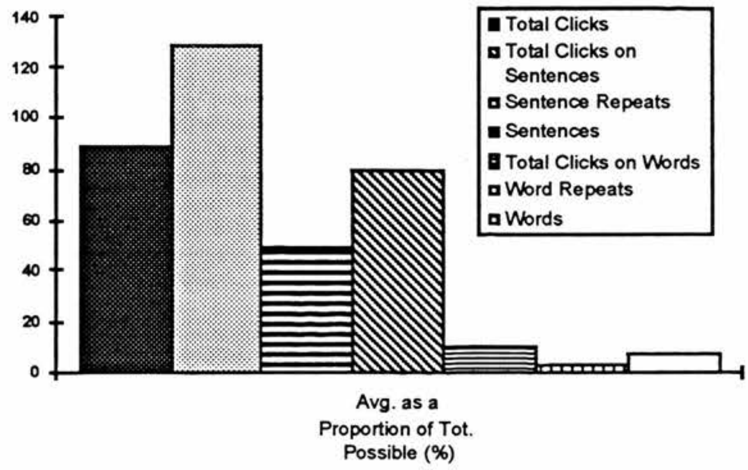
The contrast between sentence listening and word listening is clearly demonstrated in this graph. It is also evident that repeated clicking occurred even though students did not click on every possible sentence or word.

4.2.3.3 Number Clicked on Compared with Number Possible

These data can be compared with the number of possible clicks on words and sentences (e.g., if students had clicked on every sentence once, how would that compare with what they actually did?). Figure 4.2-4 summarises the results:

Figure 4.2-4

Mean Number of Different Types of Click as a Proportion of Total Possible



Again, this graph does not show us how much of the actual text was heard by the user. However, by subtracting the total number of repeats in each category from the total number of clicks, it is possible to determine how much of the possible text was listened to. Repeated clicks are included in the *Total Clicks* categories and thus the number is more than 100%.

4.2.4 Correlations Between Test Scores and Listening

In Chapter 3, Hypothesis 2 read:

> There will be a relationship between the amount of listening a student does and his/her post-test score.

A Pearson correlation '*r*'was performed on the test scores and the amount/type of mouse use. In the following table, the *r* value denotes the strength of correlation between the type of click in the columns and the post-test scores:

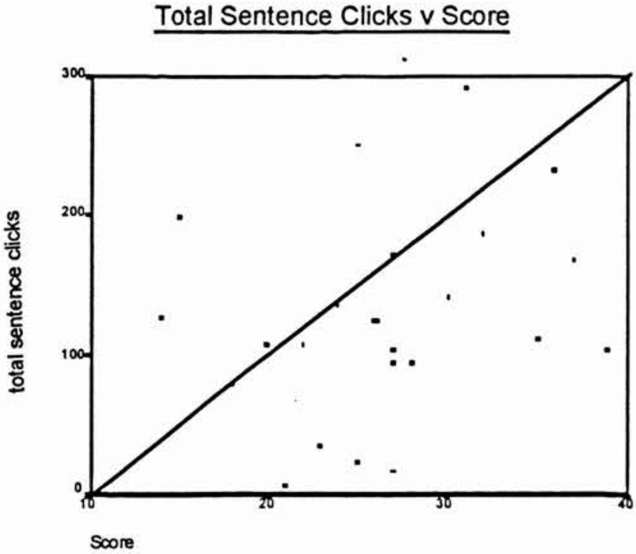
Table 4.2-5

Pearson Correlations of Score *v* Different Types of Click

		Total Mouse Use	proportion of sentences heard	proportion of words heard	number of repeated sentences	number of repeated words
Post-test Score	*r*	.23	.22	-.17	.20	-.08
	N	26	26	26	26	26
	Sig level	.26	.57	.57	.32	.69

The data in Table 4.2-5 demonstrate that there was a slight positive correlation between the total mouse use and post-test scores. There is a similarly weak correlation between the proportion of sentences clicked on, the number of repeated listening to sentences and students' post-test results. The word listening showed almost no correlation in any category to post-test scores. This is due to the limited use of the facility overall. The next graph shows the type of correlation between the total number of clicks (including repeats) and post-test scores:

Figure 4.2-5



It is clear from this graph shows that there is only a slight clustering of scores around the hypothetical direct correlation. In fact, the correlations are not significant.

Figure 4.2-6 below shows how the number of sentences to which the student listened (out of the 97 possible) relates to post-test scores. That is, repeated listening has been removed from the tally and the graph indicates the amount of the story actually heard.

Figure 4.2-6

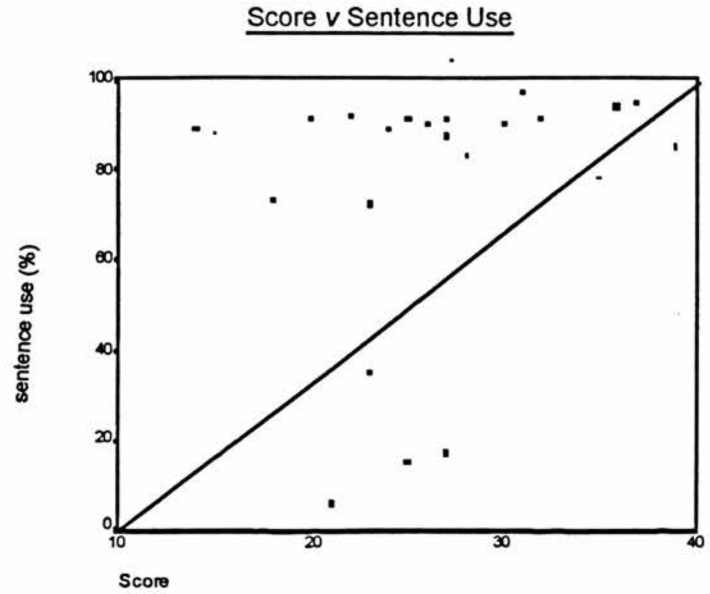
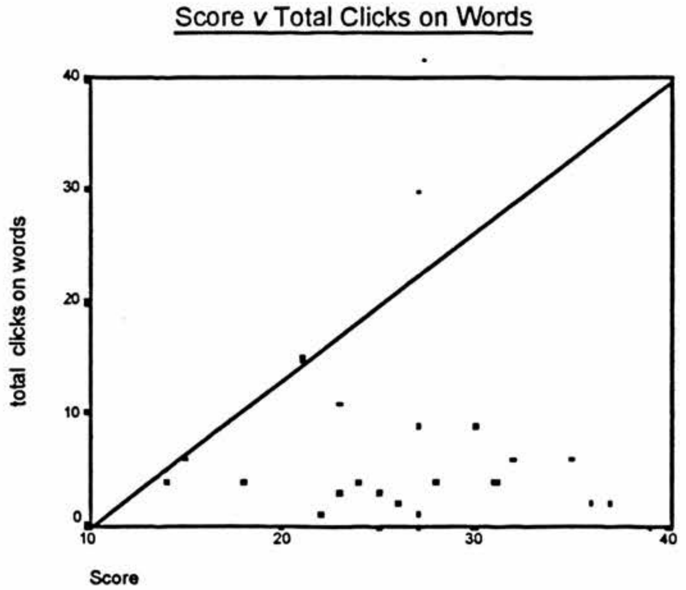


Figure 4.2-6 is perhaps more indicative of the high amount of overall sentences clicked on than the relationship between amount and score. Clearly, most readers heard a great deal of the story (on average, more than 80 out of a possible 97 sentences—about 83%).

Figure 4.2-7 shows the scores in relation to the total number of words on which the student clicked (including repeats):

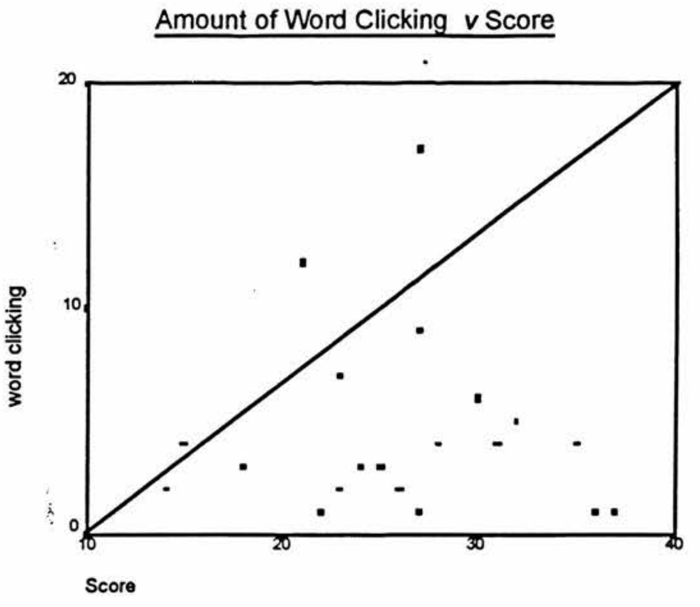
Figure 4.2-7



Again this graph shows that students avoided or missed most of the possible word listening. There is no clear correlational pattern evident in this graph.

Figure 4.2-8 demonstrates how the amount of clicking on words related to students' post-test scores:

Figure 4.2-8



Again, there is apparently little correlational patterning in these results. The most striking observation is the low amount of word level use. Considering the low correlations between overall listening and post-test scores, sentence/word listening and post-test scores, Null Hypothesis 2 cannot be rejected.

4.3 Questionnaire Hypotheses and Results

The questionnaire hypotheses will be dealt with in turn in order of presentation in the previous chapter. The first group of hypotheses concerned student ratings of difficulty and self reports of comprehension.

4.3.1 Sample Summary

43 students participated in the questionnaire part of the study—23 in the control and 20 in the experimental groups. Students were recruited in the same manner as the vocabulary test, except that they were told they would be answering a questionnaire after reading on the computer. Students were randomly assigned to

either group (see Appendix 3.4 for questionnaire frequency results and Appendix 3.5 for all questionnaire responses).

4.3.2 Data Reduction

For all items, students were given a choice using a 5-point Likert scale. Where necessary and appropriate (for statistical purposes) the data were collapsed to 2 or 3 categories. In many cases, the expected frequencies of each cell in the resulting contingency tables was too low to allow for the use of all categories. In each section describing the statistical results below, a brief explanation of how the data were collapsed will be provided.

4.3.3 Plot and Vocabulary Difficulty

The question of difficulty (or easiness) of the story's vocabulary and content was assessed by items 5 and 6 of the questionnaire and the hypotheses were as follows:

4.3.3.1 Hypothesis 3a

There will be a significant difference between control and experimental group ratings of vocabulary difficulty.

4.3.3.2 Hypothesis 3b

There will be a significant difference between control and experimental group ratings of plot difficulty.

Item 5 addresses the question of vocabulary difficulty. It reads:

The words in the story were easy.

The five point Likert scale described in the previous chapter was collapsed to see how many students indicated agreement with item 5. Table 4.3-1 shows the collapsed data:

Table 4.3-1

Frequency of Various Responses to Item 5

	Agreement with Statement		
Variable	**Other**	**Agree**	**Row Total**
Without Sound	15	5	20
With Sound	19	4	23
Column total	34	9	43

These data show that in the control group, 75% of the students did not agree that the words in the story were easy, and in the experimental group 83% did not think the words in the story were easy. A Chi-Square Distribution demonstrates that the difference between the groups is not significant. Table 4.3-2 shows the details of the test:

Table 4.3-2

Results of Chi-Square Distribution for Sound and Lexical Difficulty (Item 5)

χ^2	**value**	**df**	**Significance level**
Pearson	.37422	1	.54

If the data are collapsed so that both *disagree* options form one category and the remaining three categories form the other, the contingency table looks like this:

Table 4.3-3

Frequency of Various Responses to Item 5

	Disagreement with Statement		
Variable	**Other**	**Disagree**	**Row Total**
Without Sound	9	11	20
With Sound	15	8	23
Column total	24	19	43

Table 4.3-3 shows that 55% of the control group disagreed with this statement and about 35% of the experimental group disagreed. The following table shows the results of the Chi-Square Distribution:

Table 4.3-4

Chi-Square Distribution for Item 5 (collapsed data)

χ^2	**value**	**df**	**Significance level**
Pearson	1.77	1	.18

The difference is not significant and Null hypothesis 3a can not be rejected.

Item 6 examined the question of plot difficulty. Students agreed or disagreed with:

The plot of the story was easy.

In the control group, 50% of the respondents did not agree with this statement and the experimental group, 70% did not agree. Table 4.3-5 demonstrates the collapsed data from this questionnaire item:

Table 4.3-5

Frequency of Various Responses to Item 6

	Agreement with Statement		
Variable	**Other**	**Agree**	**Row Total**
Without Sound	10	10	20
With Sound	16	7	23
Column total	26	17	43

Again, the difference between groups is not significant. The following table shows the results of the χ^2 procedure:

Table 4.3-6

Results of Chi-Square Distribution for Sound and Plot Difficulty (Item 6)

χ^2	value	df	Significance level
Pearson	1.71306	1	.19

If the data are collapsed the other way so that both *disagree* categories are collapsed and compared against the other three categories, the following contingency table results:

Table 4.3-7

Frequency of Various Responses to Item 6

	Disagreement with Statement		
Variable	**Other**	**Disagree**	**Row Total**
Without Sound	5	15	20
With Sound	9	14	23
Column total	14	29	43

In the control group, 75% disagreed with item 6 and in the experimental group, approximately 61% disagreed. The next table shows the results of the Chi-Square Distribution:

Table 4.3-7b

Chi Square Distribution for Item 6

χ^2	value	df	Significance level
Pearson	.97	1	.32

The difference is not significant and Null Hypothesis 3b can not be rejected.

4.3.4 Enjoyment Between Groups: Hypothesis 4

Hypothesis 4 stated that :

There will be a significant difference between control and experimental group ratings of program enjoyment.

Again, the five point Likert scale described in Chapter 3 was broken into two categories. The two categories indicating agreement were collapsed into one category, and the two levels of *disagree* and *don't know* were collapsed into the other category.

Item 9 addressed the question of sound and enjoyment. It reads:

I enjoyed using the computer to read the story.

The following contingency table gives the number of responses in the collapsed categories:

Table 4.3-8

Frequency of Responses to Item 9

	Agreement with Statement		
Variable	Other	Agree	Total
Without Sound	10	10	20
With Sound	4	19	23
Column total	14	29	43

In the control group, 50% agreed with item 9. In the experimental group, 83% agreed that they enjoyed using the computer to read the story. The difference between groups is significant ($\chi^2 = 5.2$) and the following table gives the details of the test.

Table 4.3-9

Chi-Square Distribution for Item 9

χ^2	value	df	Significance level
Pearson	5.18	1	.02

Null Hypothesis 4 can therefore be rejected.

4.3.5 Plot and Vocabulary Comprehension

4.3.5.1 Hypothesis 5a

Hypothesis 5a can be stated as follows:

There will be a significant difference between control and experimental group ratings of plot comprehension.

Items 19 and 20 asked the students to rate their own comprehension of the story's plot and its vocabulary respectively. Item 19 reads:

How much of the story's plot did you understand?

The 5-point Likert scale differed from the other questionnaire items in that students were asked to estimate their comprehension as a percentage. The way in which the responses were distributed suggested that the cells should be collapsed (see Appendix 3.6 for tables of uncollapsed data for these items). The data were therefore collapsed in a similar fashion as with previous items. It was decided that a satisfactory cut-off could be made at 50% claimed comprehension. Table 4.3-10 shows these collapsed data:

Table 4.3-10

Frequency of Various Responses to Item 19

	Comprehension		
	0-50%	**50-100%**	**Row Total**
Without Sound	6	14	20
With Sound	14	9	23
Column total	20	23	43

The table shows that in the control group, 70% of the students felt that they understood more than half of the story's plot. In the experimental group, on the other hand, only 39% indicated more than 50% comprehension. With these collapsed data, the difference is significant ($\chi^2 = 4.1$). The next table shows the details of the Chi-Square Distribution on item 19.

Table 4.3-11

Results of Chi-Square Distribution on Item 19

χ^2	**value**	**df**	**Significance level**
Pearson	4.1	1	.04

With these collapsed data, Null Hypothesis 5a can therefore be rejected.

4.3.5.2 Hypothesis 5b

Hypothesis 5b made another statement about comprehension:

There will be a significant difference between control and experimental group ratings of vocabulary comprehension.

Item 20 asks the students to rate their comprehension of the story's vocabulary. It reads:

How much of the vocabulary did you understand?

If the data are collapsed in the same manner as item 19, they appear as in Table 4.3-12:

Table 4.3-12

Frequency of Various Responses to Item 20 Using Collapsed data

	Comprehension		
	0-50%	**50-100%**	**Row Total**
Without Sound	8	12	20
With Sound	13	10	23
Column total	21	22	43

In the control group, 60% of the students responded that they comprehended more than 50% of the vocabulary in the story. This can be compared with the experimental group where 43% reckoned they understood more than 50%. In this case, the difference is not significant as the Table 4.3-13 demonstrates:

Table 4.3-13

Results of Chi-Square Distribution Test on Item 20

$\chi 2$	value	df	Significance level
Pearson	1.17	1	.28

Using these collapsed data, Null Hypothesis 5b cannot be rejected.

4.3.6 Mouse Difficulty: Hypothesis 6

Hypothesis 6 stated that:

There will be a significant difference between control and treatment group ratings of mouse difficulty.

To test Null Hypothesis 6, item 3 in the questionnaire asked the students to agree or disagree with the following statement:

Using the mouse was difficult.

The next table shows the response frequencies with the data collapsed so that *agree* makes up one category and *disagree* and *don't know* made up the other.

Table 4.3-14

Frequency of Various Responses to Item 3

	Agreement with Statement		
Variable	**Other**	**Agree**	**Row Total**
Without Sound	15	5	20
With Sound	15	8	23
Total	30	13	43

In the control group 25% agreed that using the mouse was difficult compared with 35% in the experimental group. The difference is not significant as Table 4.3-15 demonstrates:

Table 4.3-15

Results of Chi-Square Distribution on Sound and Mouse Difficulty

χ^2	value	df	Significance level
Pearson	.49	1	.49

Collapsing the data in this way shows the level of agreement with the question, but it does not show us if the students actually thought it was *easy* to manipulate the mouse (as we imagined at design time). If we collapse the data into *disagree* and *other answers*, we get the following results:

Table 4.3-16

Frequency of Disagreement with Item 3

	Disagreement with Statement		
Variable	**Other**	**Disagree**	**Row Total**
Without Sound	6	14	20
With Sound	13	10	23
Total	19	24	43

In the control group 70% of the students disagreed with the notion that the mouse was difficult and in the experimental group, only 45% disagreed. However, the difference is not significant at the .05 level as Table 4.3-17 demonstrates:

Table 4.3-17

Results of Chi-Square Distribution on Sound and Mouse Difficulty: Disagreement

χ^2	value	df	Significance level
Pearson	3.05	1	.08

Thus, with the data collapsed in both directions we cannot reject Null Hypothesis 6.

4.3.7 Reading Medium Preference: Hypothesis 7

Hypothesis 7 states that,

> There will be a significant difference between control and experimental group preferences for the reading medium.

Students' preference for the way in which they read materials is addressed in item 10 of the questionnaire. It asks students to agree or disagree with the statement:

> *I would prefer to read the story in a normal book.*

The data were again collapsed to *agree* and *other* categories and the following contingency table resulted:

Table 4.3-18

Frequency of Agreement Responses to Item 10

	Agreement with Statement		
Variable	**Other**	**Agree**	**Row Total**
Without Sound	12	8	20
With Sound	11	12	23
Column total	23	20	43

Of the two groups, 40% of the control group students agreed with this statement and 52% of the experimental group students agreed. The difference between the groups is not significant as we can see in Table 4.3-19:

Table 4.3-19

Results of Chi-Square Distribution for Sound and Medium Preference (Item 10)

χ^2	value	df	Significance level
Pearson	.64	1	.42

If we reverse the direction in which the data are collapsed, so that *disagree* is one category and *others* is comprised of the *agree* and *don't know* categories, the following contingency table results:

Table 4.3-20

Frequency of Disagreement Responses to Item 10

	Disagreement with Statement		
Variable	**Other**	**Disagree**	**Row Total**
Without Sound	15	5	20
With Sound	21	2	23
Column total	36	7	43

Table 4.3-20 demonstrates that a large proportion of both groups do not prefer the computer as a reading medium. Only 25% of the control group students and 9% of the experimental group disagreed with the statement that reading a normal book was preferable. However, the difference *between* groups is not significant. Table 4.3-21 table shows the details of the Chi-Square Distribution:

Table 4.3-21

Results of Chi-Square Distribution for Sound and Medium Preference (Item 10)

χ^2	value	df	Significance level
Pearson	2.1	1	.15

Since the difference between groups is insignificant with the data collapsed in both these ways, Null Hypothesis 7 cannot be rejected.

4.3.8 Desire to Use the Computer Again: Hypothesis 8

Hypothesis 8 states that,

> There will be a significant difference between control and experimental group desire to use similar materials again.

Item 11 asked the students to agree or disagree with the statement that,

> *I would like to use the computer for reading again.*

Collapsing the data into *agree* and *other* leaves us with the following table:

Table 4.3-22

Frequency of Agreement Responses to Item 11

	Agreement with Statement		
Variable	**Other**	**Agree**	**Row Total**
Without Sound	3	17	20
With Sound	4	19	23
Column total	7	36	43

Table 4.3-22 shows that 85% of control group students agreed with this statement and 83% of experimental groups students agreed. However, the difference between the groups is not significant as the next table demonstrates:

Table 4.3-23

Results of Chi-Square Distribution for Sound and Medium Preference (Item 11)

χ^2	value	df	Significance level
Pearson	.05	1	.83

When the data are collapsed to *disagree* and *other* we see that there was very little disagreement with this statement, as shown in table 4.3-24.

Table 4.3-24

Frequency of Disagreement Responses to Item 11

	Disagreement with Statement		
Variable	**Other**	**Disagree**	**Row Total**
Without Sound	19	1	20
With Sound	20	3	23
Column total	39	4	43

We can see that a very small number of students (5% and 13% in the control and treatment groups respectively) disagreed with the statement in item 11. However, the difference between groups was not significant.

Table 4.3-25

Results of Chi-Square Distribution for Sound and Medium Preference (Item 11)

χ^2	value	df	Significance level
Pearson	.82	1	.37

Since there is no significant difference between groups with the data collapsed into *agree/don't know* categories and then *disagree/don't know* categories, Null Hypothesis 8 cannot be rejected.

4.3.9 Computer for Language Learning: Hypothesis 9

In Hypothesis 9 we posited the following:

> There will be a significant difference between control and experimental group perception of the computer as a language learning tool.

To determine students' perception of the computer as a language learning tool, we asked them to agree or disagree with item 12 which was as follows,

> *Computers are good for learning languages.*

Table 4.3-26 shows the collapsed data:

Table 4.3-26

Frequency of Agreement Responses to Item 12

	Agreement with Statement		
Variable	**Other**	**Agree**	**Row Total**
Without Sound	6	14	20
With Sound	3	20	23
Column total	9	34	43

These data show us that 70% of the control group and 87% of the experimental group agreed with this statement. The Chi-Square Distribution below in Table 4.3-27 shows that the difference is not significant.

Table 4.3-27

Results of Chi-Square Distribution for Sound and Medium Preference (Item 12)

χ^2	value	df	Significance level
Pearson	1.86	1	.17

Again, the data are collapsed in the other direction as seen in Table 4.3-28:

Table 4.3-28

Frequency of Disagreement Responses to Item 12

	Disagreement with Statement		
Variable	**Other**	**Disagree**	**Row Total**
Without Sound	19	1	20
With Sound	23	0	23
Column total	42	1	

The data in this table show us that very few students in either group disagreed with the notion that computers are good for learning languages but the difference between groups is not significant. Table 4.2-29 gives the details of the χ^2 procedure:

Table 4.3-29

Results of Chi-Square Distribution for Sound and Medium Preference (Item 12)

χ^2	value	df	Significance level
Pearson	1.18	1	.28

4.3.10 Predispositions

Hypotheses 10 through 13 and hypothesis 18 were designed to rule out the possibility that students' predispositions towards the computer would be correlated to their enjoyment of the experimental program. Of these hypotheses, numbers 11a and 11b apply specifically to the students who were listening. For these items, the statistical analysis was performed on the experimental group only. Each of the null hypotheses will be listed in the next sections by number.

4.3.10.1 Hypothesis 10

Hypothesis 10 surmised that students who like computers would like the reading experience. The hypothesis is as follows,

> There will be a significant correlation between how much students like computers and how much they liked the computer-based reading exercise.

In the overall group, there was a strong correlation between how much students like computers and how much they like using the computer to read the story. The following table summarises these data:

Table 4.3-30

Correlation Between Enjoyment of Computers and Enjoyment of the Story

	Correlation	N	Significance
Combined Group	.41	43	.007

These results demonstrate a significant correlation between responses to the item *I like computers* and the item *I enjoyed using computers to read the story*. Null Hypothesis 10 can be rejected.

4.3.10.2 Hypothesis 11a

In Hypotheses 11a we surmised that,

> There will be a correlation between students' enjoyment of listening while reading and their enjoyment of the computer for the experimental task.

Items 8 and 9 are as follows,

Item 8:

> *I like listening to stories while reading.*

Item 9:

> *I enjoyed using the computer to read the story.*

Table 4.3-31 summarises the Spearman correlation.

Table 4.3-31

Correlation Between Items 8 and 9

	Correlation	N	Significance
Experimental Group	.20	23	.36

It appears that there is very little relationship in the experimental group between students' general enjoyment of reading and listening and their enjoyment of the computer-based reading. Null Hypothesis 11a cannot be rejected.

4.3.10.3 Hypothesis 11b

Hypothesis 11b stated that,

> There will be a correlation between students' enjoyment of listening while reading with a teacher and their enjoyment of the computer task.

It was thought that the enjoyment of the reading might be influenced by students' predisposition to reading with a teacher. To rule out this possibility, item 15 was included in the questionnaire. Item 9 asked students to agree or disagree with:

I enjoyed using the computer to read the story.

and item 15 was as follows:

I enjoy reading and listening with a teacher.

The following table shows a small and insignificant correlation between students' predilection for teacher-based reading and listening and the type of reading and listening they encountered in the experiment.

Table 4.3-32

Correlation Between Items 9 and 15

	Correlation	N	Significance
Experimental Group	.18	23	.40

As Table 4.3-32 demonstrates, the correlation was also weak and insignificant. We therefore cannot reject Null Hypothesis 11b.

4.3.10.4 Hypothesis 12

Hypothesis 12 was also designed to examine the relationship between reading enjoyment and program enjoyment. The hypothesis was as follows:

> There will be a correlation between students' enjoyment of short stories in Japanese and their enjoyment of the computer-based reading.

The responses from item 9 were correlated to item 18 which read,

I enjoy reading in my own language.

There was a weak negative correlation between students' enjoyment of reading in Japanese and their enjoyment of the computer-based reading as seen in Table 4.3-33.

Table 4.3-33

Correlation Between Items 9 and 18

	Correlation	N	Significance
Combined Group	-.23	43	.14

The correlation is not significant so we cannot reject Null Hypothesis 12.

4.3.10.5 Hypothesis 13

Hypothesis 13 proposed a similar notion to hypothesis 12 but this time the question was whether enjoyment of reading in English was related to enjoyment of the experimental reading. It states that,

> There will be a correlation between students enjoyment of short stories in English and their enjoyment of the computer-based reading.

In this case, the hypothesis predicts that items 9 and 17 will correlate. Item 17 is as follows:

I enjoy reading short stories in English.

Table 4.3-34 shows no correlation between students' enjoyment of general English reading and reading on the computer in the combined group:

Table 4.3-34

Correlation Between Items 9 and 17

	Correlation	N	Significance
Combined Group	-.05	43	.73

Null Hypothesis 13 cannot be rejected.

4.3.10.6 Hypothesis 14

Hypothesis 14 suggests that there is a correlation between general problems with reading and problems with the vocabulary and plot of the experimental story. The hypothesis is as follows:

> There will be a correlation between students' difficulty with reading in English in general and their difficulty with the computer-based reading.

A correlation matrix was established between items 5, 6, and 16. The items were:

Item 5

The words in the story were difficult.

Item 6

The plot of the story was difficult.

Item 16

Reading in English is difficult.

Again, the data show a very weak correlation between students' perception of reading as difficult and their perception of the plot and vocabulary difficulty in the experimental story. This table shows a correlation matrix between the three questionnaire items in the combined group.

Table 4.3-35

Correlation Matrix: items 5, 6, and 16

	item 5	item 6
Item 6 correlation	.50	
significance	.001	
item 16 correlation	-.09	-.07
significance	.56	.67

As might be expected, items 5 and 6 correlate strongly. However, there seems to be no relationship between students' feelings that reading is difficult and their ratings of vocabulary and plot difficulty in the experimental text. Null Hypothesis 14 cannot be rejected.

4.3.11 Computer Usefulness: Hypothesis 15

Hypothesis 15 was designed to examine the students' opinion of the experimental reading in terms of its usefulness.

> There will be a significantly different response between groups that the computer-based reading is useful.

Students were asked to agree or disagree with item 14 which read,

> *Using the computer for reading is a waste of time.*

The data were collapsed in the same fashion as previous items. The next table summarised the *agree* and *other* categories first:

Table 4.3-36

Frequency of 'agree' and 'other' Responses to Item 14

	Agreement with Statement		
Variable	**Other**	**Agree**	**Row Total**
Without Sound	19	1	20
With Sound	23	0	23
Column total	42	1	43

Clearly a large proportion of students in both groups did not agree with the proposition made by item 14. In the control group, 95% of students did not agree and fully 100% of the experimental group did not agree. The difference is not significant as the following table demonstrates.

Table 4.3-37

Results of Chi-Square Distribution for 'agree' and 'other' Responses to Item 14

χ^2	value	df	Significance level
Pearson	.16	1	.69

Repeating the procedure for *disagree* and *other* categories, the following contingency table results:

Table 4.3-38

Frequency of 'disagree' and 'other' Responses to Item 14

	Disagreement with Statement		
Variable	**Other**	**Disagree**	**Row Total**
Without Sound	2	18	20
With Sound	1	22	23
Column total	3	40	43

Table 4.3-38 shows that most of the students in fact disagreed with item 14 (90% in the control and 96% in the experimental group). Again, the difference between groups is not significant as seen in Table 4.3-39:

Table 4.3-39

Results of Chi-Square Distribution for 'disagree' and 'other' Responses to Item 14

χ^2	value	df	Significance level
Pearson	.53	1	.47

Using these data from both tables, Null Hypothesis 15 cannot be rejected.

4.3.12 Experience and Enjoyment

The items querying the students' previous computer experience (numbers 1 and 2) were designed to give the researcher an indication of how students' familiarity with computing related to their enjoyment of the exercise and difficulty with the mouse.

4.3.12.1 Hypothesis 16

Hypothesis 16 stated that,

> There will be a significant correlation between previous computer experience and enjoyment.

In the tables to follow, correlations are read at the intersecting cell of each questionnaire item. Items 1 and 2 (computer and mouse experience respectively) were correlated with item 9:

Table 4.3-40

Strength of Correlation Between Experience and Enjoyment (Items 1, 2, 9)

	Item 1	Item 2
Correlation Item 9	-.16	-.29
significance	.32	.05

As the table demonstrates, there was no significant correlation between previous computer and mouse experience and enjoyment. There is a moderate negative correlation between mouse experience and enjoyment (the more inexperienced students were, the less they agreed that they enjoyed it). Nevertheless, as stated, Null Hypothesis 16 cannot be rejected.

4.3.12.2 Hypothesis 17

Hypothesis 17 states:

> There will be a significant correlation between previous computer experience and interface difficulty.

Item 3 asked the students to agree or disagree with:

Using the mouse was difficult

As has been said, items 1 and 2 examined the students' computer and mouse experience respectively. The hypothesis includes both types of experience. The correlation between these three items is shown with corresponding levels of significance in the following table:

Table 4.3-41

Strength of Correlation between Experience and Mouse Difficulty (Items 1, 2, and item 3)

	Item 1	Item 2
Correlation Item 3	-.07	-.18
significance	.63	.24

There are no significant correlations in Table 4.3-41. Null Hypothesis 17 can not be rejected.

4.4 Other Data

Several trends appeared in the data that were not predicted in the previous chapter. These will be covered in the following paragraphs.

4.4.1 The Computer as a Reading Medium: Combined Group Results

Responses to item 10, for example, showed that the groups were slightly ambivalent about the computer as a reading medium compared with a normal book. To recap, item 10 states that:

I would prefer to read the story in a normal book.

In the combined group, there was a significant difference in the distribution of responses on the 5-point Likert scale ($\chi^2 = 17.82$, P <.05). Collapsing the data shows that the primary difference was between *disagree* and *other* categories. Only 16% of the combined group disagreed with item 10 ($\chi^2 = 19.56$). The result of the Chi-Square Distribution on the distribution of *agree* and *other* categories was not significant (see Appendix 3.7 for the details of this test). This was due largely to the number of students responding *don't' know* (16 out of 43 students or about 37%).

4.4.2 Overall Computer Experience

Contrary to our expectations, there was a significant number of students with very low levels of previous computer experience. Table 4.4-1 shows how the whole group responded to the item about computer experience:

Table 4.4-1

Results of Chi-Square Distribution test on Previous Computer Experience (Item 1)

Category	Observed	Expected	Residual
1(none)	13	8.6	4.4
2(little)	14	8.6	5.4
3(once in while)	4	8.6	-4.6
4 (often)	9	8.6	.4
5(very often)	2	8.6	5.6
	43		
DF	4		
χ^2	11.7674		
significance	.0192		

Only 26% of the students reported that they had used a computer often or very often, while 40% of the students indicated little or no previous computer experience. The data can be logically collapsed into two categories: categories 4 and 5 were considered 'experienced' and categories 1, 2, and 3 were 'inexperienced'. The difference in frequency of responses between the two categories is significant (χ^2 = 8.4, P < .01) showing that students chose to describe themselves as 'inexperienced' (see Appendix 3.8 for the details of these Chi-Square Distributions).

Students' reports of mouse experience can be compared to these data. Table 4.4-2 shows the uncollapsed data and a significant χ^2 result:

Table 4.4-2

Results of Chi-Square Distribution on Previous Mouse Experience (Item 2)

Category	Observed	Expected	Residual
1(none)	23	8.6	14.4
2(little)	14	8.6	5.4
3(once in while)	2	8.6	-6.6
4 (often)	3	8.6	-5.6
5(very often)	1	8.6	-7.6
TOTAL	43		
DF	4		
χ^2	42.93		
significance	.00		

On item 2, more than 86% of the students reported no or little mouse experience and only 9% of the students reported a high level of previous experience. If these data are also collapsed into 'experienced' and 'inexperienced' in the same ways as for item 1, the difference between categories is significant (χ^2 = 28.49, P < .01). On the whole, students responded that they were also inexperienced mouse users.

4.4.3 Previous Experience and Mouse Difficulty

One would expect that reports of mouse difficulty would be closely linked to previous mouse experience. To determine if this was the case, a Spearman correlation was performed between items 2 and 3 (mouse experience and mouse difficulty, respectively). There was a weak, insignificant positive correlation between previous mouse experience and mouse difficulty (.17 P >.05).

4.4.4 Previous Experience and Story Difficulty (story plot and story vocabulary)

Interestingly, the data suggest that there was a relationship between previous experience and ratings of story difficulty.

The questions involving story difficulty were items 5 and 6. We read:

Item 5: *The vocabulary was easy.*

Item 6: *The plot of the story was easy.*

Using Spearman correlation in the experimental groups shows that there is a significant correlation between previous computer experience and ratings of plot/vocabulary difficulty and also a strong correlation between previous mouse experience and vocabulary difficulty. Table 4.4-3 summarises these results:

Table 4.4-3

Strength of Correlation Between Experience and Difficulty (Items 5 & 6 v Items 1 &2) in Control and Experimental Groups

		Control		Experimental	
		Item 5	Item 6	Item 5	Item 6
Spearman	Item 1	.05	.05	.46	.36
Sig.	Item 1	.85	.83	.03	.09
Spearman	Item 2	-.12	.06	.20	.41
Sig.	Item 2	.62	.82	.35	.05

It should be noted that in the control group, there is no significant correlation between any of the items in the above matrix.

4.4.5 Items 5 and 6: Difficulty in the Combined Group

When analysing the data for Hypothesis 3a and 3b, it was noticed that a large number of students in both groups chose to respond in the *disagree* categories on the items to do with vocabulary and plot difficulty. To determine whether the difference in distribution between categories was significant, a Chi-Square Distribution was performed on the combined groups and with all categories (see Appendix 3.10 for the details of this test). On item 5, across all categories in the combined group, it was found that the distribution of observed responses was indeed significantly different from the expected responses (χ^2 = 37.81, $P < .05$). Similarly, on item 6, the Chi-Square result was significant (χ^2 = 20.37, $P < .05$).

To determine in which categories the responses differed significantly, the categories were collapsed so that *agree* was combined and compared against the others, *disagree* was combined and compared against the others, and *don't know* was compared against both *agree* and *disagree*. For item 5, the relevant significant difference was found to be between *agree* and *all others* with more responses in the *all others* categories. That is, more students chose not to agree that the vocabulary was easy than to agree. For item 6, the relevant significant difference was found to be between *Don't Know* and *All Others.* In this case, significantly more students chose not to disagree with the statement that the plot was easy.

4.4.6 Item 14: Computer Usefulness

To the item suggesting that using the computer for reading was a waste of time, students responded overwhelmingly in the *disagree* categories. When the *disagree* categories are combined and compared against the other three categories, we see that approximately 96% of the students in the combined group disagreed or strongly disagreed with this statement. A Chi-Square Distribution shows that these results are significant (χ^2 = 31.84, $P < .01$).

4.5 Summary

The data in this chapter are derived from a sample of 102 students. 59 of these participated in a computer-based reading activity followed by a 48 item multiple choice vocabulary post-test and 43 did the same computer activity followed by a 20 item questionnaire. In both the post-test and the questionnaire sample, students were divided into a control group and a treatment group. The post-test data were analysed using an independent samples t-test. Further analysis of the type of computer use was done using a Pearson correlation. For analysis of the non-parametric questionnaire data, the researcher used Spearman correlations, and both a two sample and a single sample Chi-Square Distribution. For all tests, a significance level (α) of .05 was assumed.

Results from they vocabulary post-test do not permit rejection of Null Hypotheses 1 and 2. The data from the log records revealed that students used a large majority of the sentence listening opportunities and a very low proportion of the word listening possibilities.

The questionnaire data showed that students in the experimental group enjoyed the reading more than those in the control group and that the difference was statistically significant. The data also suggested that the two groups differed in the amount of the plot they claimed to comprehend. There were significantly more students in the control group who claimed to comprehend more than 50% than in the experimental group. On the other hand, there was no difference between groups for the number of students who claimed to understand more than 50% of the vocabulary. With regard to vocabulary and plot difficulty, medium preference, and mouse difficulty, there was no difference between the two groups. There appeared to be no relationship between enjoyment of the computer-based reading exercise and enjoyment of:

reading while listening (in the experimental group);
reading while listening with a teacher (experimental group);
reading in Japanese;
reading in English.

Similarly, there was no significant correlation between:

previous experience and enjoyment of the computer-based reading;
previous computer experience and difficulty with the mouse.

There was a significant correlation between how much students liked computers and their enjoyment of the computer-based reading.

Table 4.5-1 summarises the hypotheses and results:

Table 4.5-1

Summary of Null Hypotheses and Results

Null Hypothesis	Brief Description	Reject?
1	no difference between groups on post-test means	N
2	no relationship between use and post-test score	N
3a	no difference between group for difficulty of vocabulary	N
3b	no difference between groups for difficulty of plot	N
4	no difference between groups for program enjoyment	Y
5a	no difference between groups for plot comprehension	Y
5b	no difference between groups for vocabulary comprehension	N
6	no difference between groups for mouse difficulty	N
7	no difference between groups for reading medium preference	N
8	no difference between groups' desire to use the program again	N
9	no difference between groups for their opinion of the computer as a language learning tool.	N
10	no relationship between enjoyment of computers and enjoyment of the computer-based reading	Y
11a	no relationship between students' liking listening and reading and enjoyment of experiment	N
11b	no relationship between students' like of listening and reading with a teacher and enjoyment of experiment	N
12	no relationship between students' enjoyment of stories in Japanese and their enjoyment of experiment	N
13	no relationship between students' enjoyment of stories in English and their enjoyment of experiment	N
14	no relationship between general English reading difficulty and difficulty of reading in experiment	N
15	no difference between groups' perception of computer usefulness	N
16	no relationship between computer previous experience and enjoyment	N
17	no relationship between previous computer experience and difficulty of experiment	N

In addition to these results, the data showed that students in the combined group did not prefer the computer to a book as a reading medium, nor did they see the exercise as a waste of time. Both groups also had very little previous computer or mouse experience.

Other statistical analyses revealed a correlation between previous mouse and computer experience and ratings of story difficulty in the experimental group and also a strong correlation between previous mouse experience and plot difficulty.

Table of Contents: Chapter 5

CHAPTER 5
DISCUSSION

Introduction

Chapter 5 is divided into three sections: (1) a discussion of the post-test; (2) a discussion of the questionnaire results with an emphasis on the quantitative data; and (3) a discussion of the open-ended questionnaire responses paying specific attention to improvements in the materials' design. Throughout each of these sections, data from the open-ended questionnaire items will be used to provide insight into various quantitative results. Section 1 will discuss the post-test results in terms of hypotheses 1 and 2. We will examine possible reasons for the lack of measurable differences between post-test groups and suggest design improvements that might be made in future experiments of this type. This section will also discuss the weak correlations between the amount of listening and post-test scores. Section 2 will then discuss the questionnaire data with regard to difficulty, perceived comprehension (of the story and vocabulary), enjoyment, and the preferred reading medium. Section 3 of Chapter 5 will look at the open-ended responses from the questionnaire and their implications for the reading exercise as classroom materials. We will also discuss students' expectations of hypermedia materials, and changes that might be made to the program to improve its pedagogic value.

5.1 Post-Test Results

5.1.1 Group Comparison

Hypothesis 1 stated that there would be a significant difference between the post-test means of the control and experimental groups. The data from the previous chapter shows that Null Hypothesis 1 cannot be rejected. This section will focus on the reasons for this lack of differentiation, discussing how the effects of reading while listening might be better measured, and also how a greater positive effect might be achieved. Factors that will be considered are: exposure time, text length, computer novelty, negative experimental effects, learner level, inadequate listening, and the possibility of no effect.

5.1.1.1 Abbreviations

In this and the following sections, reference will be made to students' responses to the open-ended items on the questionnaire. Each student has been numerically coded by number, group, and sex so that the reader can compare the summaries made below with the actual statements made by students. The abbreviation 'SN:' has been used to denote 'student number' or 'student numbers'. For example, 'SN: 23C' would mean 'student number 23, control group'.

5.1.1.2 Time

There are several possible explanations for the lack of measured effect in hypothesis 1 of this thesis. The first of these is time. We have already discussed in Chapter 2 that incidental acquisition of vocabulary during normal reading proceeds in small increments (Hulstijn 1992, Nation and Coady 1988). It may be equally true of incidental acquisition through reading and listening that if there *is* an effect, it is too small to be measured after one exposure. In the present design, students used the materials over a brief period (30 minutes), and only once. The exercise was not aimed at measuring improvements in students' lexicon over the long-term. However, it is likely that students will not remember both the new words and the contexts in which they appeared in one short exercise. Just as extensive reading is seen as a source of new vocabulary, extensive reading and listening might also produce vocabulary growth. The possibility of significant differences might have been improved had the students used this kind of system two or three times per week over several months.

Long-term and frequent exposure to new words in context encourages growth of the mental lexicon. It is also possible that a longer period of use would have overcome the problem of unfamiliarity with the computer and the system. There are, however, two restrictions on a long-term experiment. First, while we might have seen a more pronounced effect of listening on vocabulary acquisition over a longer period, it would be extremely difficult to know whether improvements were made because of the listening procedure or because of other variables for which we could not easily account. The second restriction was practical. When the project began, the total disk space on the available computer was 80 megabytes. The text, with sound and program files, used more than half of that space (43 megabytes) and the computer's system files (including Windows, DOS, Guide, and Word for Windows) used most of the remaining resources. It would have been impossible, with the available resources, to create the number of texts that would be required by students over weeks or months. Indeed, the experiment was designed partially with the justification of future larger systems in mind. One further constraint was that the students in the target population were only at the Institute for a three week period. Most of the students' first week was devoted to orientation; at best, only two weeks would have been available for a 'long-term' study. Clearly, extensive exposure would have been impossible in this time frame.

5.1.1.3 Text Length: Word Frequency and Context

The likelihood of significant differences betwee groups might have been further enhanced had students been asked to read a longer text where target words appeared in several contexts and/or more than once. This possibility was appreciated at the design stage but again the length of the text was severely limited by the size of the hard disks. Furthermore, since the materials' authenticity was a primary concern at the development stage, it was felt that any repetition of target words in a short text (unless it happened naturally) would be artificial and therefore unsatisfactory. In keeping with the idea of authentic texts was the notion that the reading activity should be one that might be performed in a normal self-access session at the institution where it took place.

Context richness was another area that might be controlled but was also constrained by text authenticity and text length (the words might have appeared in a more varied context had the text been longer). Other researchers (e.g., Mondria and Witt-de Boer 1991) chose to vary contexts by varying the words surrounding the target word; we chose to keep the written context the same and vary the way in which the whole (authentic) text was encountered. Again, the idea was, as much as possible, to mimic a self-access session that we could envision ourselves conducting with 'real' students. It is becoming a common complaint that research in reading and vocabulary acquisition uses unnatural or unlikely texts in unnatural or unlikely teaching materials (see Bernhardt 1991 for a discussion of this topic). This researcher still maintains that while natural texts may not give the necessary stimuli to produce statistically significant results, experimentation is more valid if authentic texts are used. Significant results using an authentic text would have been encouraging. However, the strength of the results in the present study may have suffered by using target words with variable and uncontrolled contexts. In Chapter 2 we argued that increased acoustic information might enhance students' memory of the context and therefore increase their ability to recognise the meanings of new words later on in the test. This idea is summarised by Mondria and Wit de Boer (1991: 251, citing Schouten-Van Parreren 1985)

> ...it became clear from her [Schouten-Van Parreren] research that test-subjects remembered the meaning of words because the word-form called to mind a brief combination of words, usually a combination that they had come across a number of times already, and/or very pregnant combinations.

In other words, it is probably more beneficial to encounter a word in context than in citation form. It follows, we argued, that an enhanced context (as listening is in this experiment) would enhance the acquisition of new words. We also stated that listening to the words might trigger the memory of previous encounters with those words if knowledge (as defined in Chapter 2) was hazy.

It is generally agreed that the 'richer' the context, the more likely readers are to acquire new words (see Mondria and Wit-De Boer 1991 for what makes a context rich). Future investigation may prove that the addition of sound enhances the memory of the context but the data from this experiment do not suggest that this is so.

It is expected that the way in which students use new materials (i.e., whether they browse, read every word, skim, scan, etc.) will be characterised by variation—both within the learner and between learners. The difficulty of characterising learners' changing personae was discussed in Chapter 1 and variation between learners was discussed in Chapter 2. Despite the recognition of patterns in multimedia use, there is still little to suggest which patterns are optimal for which learners. There is still less to suggest how to control the way in which students use the materials (or even whether it would be desirable to do so). We may surmise that part of the variation in students' use of multimedia materials is due to the fact that many students are unfamiliar with both the possibilities of the medium and the most advantageous means of using computer-based materials. This is not to say, however, that increased familiarity will lead to increased uniformity of user styles. Rather, it suggests that inexperienced users have not yet determined what is optimal for them. They have not developed a personal learning strategy for the new medium.

When dealing with any new teaching medium, it is possible that traditional learning strategies (for whatever the task may be) might be ill-suited or less than optimal for the new task. When we ask students to do things that they do not normally do—and whether that is because the task is unnatural or just because it is new—we are asking them to approach the new task with old skills. It does take time to adapt before the new strategies are learned.

Learners in the multimedia world are presented with many more difficulties than just deciding whether they should skim or scan, infer or cogitate. They need to decide not only *how* and *what* but also *with which tool*; they have many potential tools at their disposal and each one might require a different type of individual learning strategy. Some of the features available in today's multimedia materials including on-line glossaries, spoken word definitions, spoken summaries and explanations and video clips or animation accompanied by sound (one or more of which may be available at the same time) have no counterpart in traditional materials. Students now not only have to decide which features they want to use, but they have to learn how to take advantage of them. It is possible that in the process of learning just how to use these new features, there will be a great deal of trial and error and until students develop appropriate strategies, the positive effects of a new medium might be diminished. There may very well be a 'learning to learn' effect, i.e., learning rates may increase as practice progresses (Nation 1982: 17). The argument here, therefore, is that the more different the multimedia activity is from a related traditional task, the longer it will take students to develop strategies for its efficient use.

Hypertext typically has been lauded because it permits (what many claim to be) an intuitive way of reading and acquiring information. The data from this research, however, suggest that intuitiveness of both materials and interface is largely a function of familiarity and what was traditionally an intuitive and successful acquisition (or learning) strategy might actually be a hindrance when reading with sound (or any other multimedia feature). This argument is best summed up by one of the several Medical English students (i.e., a higher level than the majority of the participants) who participated in the questionnaire group. She commented that,

> The sound drew my attention to the difficult words. Normally I would have just skipped them and gone on without thinking about them but since there was sound attached to them, I got the feeling that there were many words and many other things that I did not know.

This spontaneous reflection by a student (who incidentally, also admitted to struggling with the exercise) offers us a useful insight into the potential problems with multimedia in general and talking books specifically. It is possible that the sound, and the way in which it was implemented, forced students to stop and consider more than they naturally would have. For example, if the students' natural tendency to skip the *unknown* and move unperturbed on to the *known* is actually inhibited by the addition of sound, then it is incorrect to perceive (as we predicted in Chapters 3 and discussed in the introduction to this section) the addition of sound as an enhancement. In fact, the sound (and particularly the sound attached to individual words) may be encouraging a stop-and-consider approach in students who might better utilise a skip-it-and-move-on reading strategy. We probably should not, therefore, expect any positive effects from a new method to appear until students have been given enough time to learn how to take advantage of it. Indeed, some hypermedia additions may prove detrimental in terms of vocabulary acquisition. Other researchers have noted similar findings. Kang (1995: 54), for example, in a study of vocabulary acquisition through giving computer-based contextualised instructions noted that,

> Considering that a new instructional approach such as this [computer-based context-embedded instruction] may, at first, overwhelm the learner and that the affective domain is also a very important aspect for actual use, care should be taken to minimize the overall cognitive load imposed on the learner in such a way as to help the learner get the most of the experience.

This again addresses the question of experience and its relationship to the effectiveness of the materials.

5.1.1.4 Positive Effects countered by Negative Effects (experience)

Another explanation for the lack of measured effect was that any positive effects (in terms of increased lexical knowledge) were countered by the low previous computer experience of the students. As explained in Chapter 4, it was observed about halfway through the post-test data collection stage, that the lack of previous experience was a potential problem. In fact, we were fairly certain at design time that of all the possible ways of physically interacting with a computer, using a mouse was the easiest and that even novice users could quickly gain enough dexterity to complete the present exercise without it affecting the results. We did not suspect that

Japanese students (coming from a culture of technophiles) would be unfamiliar with using the mouse. In reality, in the questionnaire group, where students were specifically asked about their previous computer experience, a significant number of students indicated a low level of computer (and mouse) familiarity. Since both groups were drawn from a very similar population, there would be little reason to expect that students in the vocabulary test groups were any more experienced than those of the questionnaire groups. Clearly it is worth re-evaluating the preconception that Japanese students are highly computer literate.

Many students did comment that the mouse was difficult in some way. Apart from the open-ended questionnaire responses listed below, one student was observed to slam the mouse into the table in frustration. Individual students in the questionnaire sample noted that,

- the slow response of the mouse was irritating;
- the operation of the mouse should be made easier;
- it took a lot of time to master the program;
- more familiarisation with the computer was needed;
- it was difficult to make the mouse respond quickly.

If this perceived difficulty translates to frustration and a sense of being 'out of control', then it would not be unreasonable to expect that this distraction would decrease the amount of attention paid to the text. We might also argue that because so many students specifically pointed out their difficulties with the mouse, the positive effects of listening were undermined by the problem of computer familiarity and perhaps the problems of a difficult interface.

The fact that previous experience correlates significantly with plot comprehension is somewhat of a mystery. The possibility exists that this is a statistical anomaly. However, it may be that novice computer users use a great deal of effort manipulating the mouse and that this distraction makes the entire exercise seem incomprehensible. Lack of experience did not necessarily cause the students to think they did not understand the story plot. However, lack of experience might very well have contributed to a general sense of unease; and frustration caused by interface difficulties may have led to a lower estimation of comprehension.

5.1.1.4.1 Suggested Design Modifications

Despite controlling for many other variables in this experiment, the results from this research show that the lack of control over previous computer experience was an oversight. It might, however, be a further error to argue that there would be no positive effect on vocabulary acquisition in more experienced users. The following suggestions might therefore be included in a follow up study to this experiment: (1) give students extended training periods over a few days preceding the experiment; or (2) select students with more previous computer/mouse experience (see section 3 in this chapter for a discussion of changes that might be made to the interface itself). Of the two options, number 1 is preferable from an experimental perspective. Given the kind of students available in the current situation, it might prove difficult to find students who are both willing to participate and also confident with a computer.

We can sum up experience with two principles that might explain the absence of the positive effects of new computer based materials: (1) the importance of overall previous computer experience (including confidence with a mouse and GUI) should not be underestimated; and (2) expertise (including the user's ability to develop a personally efficient learning style for the new type of materials) must be permitted to develop in students by giving them time to practise not just with the computer but also with the task itself. This is important because students may be tempted to give up a potentially beneficial method.

5.1.1.5 Story Choice

The Blue Bouquet is a slightly unusual story. The plot involves a man going through an attempted robbery and the thief makes the strange request for the man's eyes. Some of the students commented upon completion of the reading that they were not sure they understood the story correctly because of this. Others commented in the questionnaire that the story was beyond their level and that they should have had more story choice. Despite every effort taken beforehand to choose a story that many students would find engaging (and interviews with pilot test students confirmed that *they* thought it was interesting and comprehensible), the fact that so many students in the questionnaire group made negative comments about the story itself suggests that a simpler story (in particular, a more transparent plot) might have been more appropriate. As many researchers have argued, learning takes place only after some comprehension (Krashen 1989, Rumelhart and Norman (1975) cited in van der Linden (1993)). If the level of vocabulary and plot complexity were too difficult for the present population sample, one would expect little incidental vocabulary acquisition in *either* group; and it would come as no surprise, that there would be no difference *between* groups on the post-test.

5.1.1.6 Cross section of Language Levels was not sufficient

One other explanation for the negligible difference in the mean of post-test scores is that students of different language abilities may be able to benefit more or less from the system than those in the current sample. There are two facts about the current sample that complicate this discussion: (1) we do not have data to suggest the overall language ability of all the students and (2) the scores we have show a generally low level of language proficiency.

One might expect, intuitively, that lower level students would not be able to take advantage of the acoustic information as well as higher level students. Henning (1974), as discussed in Chapter 2, found that lower level students in fact used more acoustic information than higher students (who tended to use more semantic information when storing words in the mind). This does not mean, however, that lower level students use additional acoustic information to acquire words, nor that additional acoustic information is beneficial to the acquisition process. In fact, the opposite may very well be true—lower level students may be overwhelmed by the extra acoustic information and their limited L2 ability may be further confused by difficulty accessing the sound feature. Lower level students are also less likely to understand the initial oral instructions, less likely to feel comfortable with the text itself and also less likely to understand the items and their alternatives on the post-test.

Unfortunately, from the data available in this study, we cannot determine if the effects of listening while reading differ in various levels of Japanese learners.

5.1.1.7 Low Use of a Key Feature (word-listening)

Students did not use the word-listening facility extensively. The reasons for this might be the interface design (see later sections on Materials Design) or might be due to some other reason. It is not clear if there would have been any difference in post-test scores had the students taken more advantage of this feature. The low amount of word-listening makes it difficult to assess the weak correlation between word listening and post-test scores. If there had been more variability in the degree of use, the correlation might have been more meaningful. One possible reason why students did not use the word-listening facility is that they were discouraged by the fact that most of the words were completely unknown to them. Again, the story difficulty might have contributed to this problem. Students might have felt that it was purposeless to listen to words of which they had no hope of gaining more understanding.

There is, of course, the possibility that rather than seeing the listening as an opportunity to clarify an unknown word, they might have considered it a reminder of how little they knew, i.e., students *chose* not to look for and use the word-only listening facility for their own reasons. This is in keeping with the observation of the Medical English student mentioned above in section 5.1.1.3. In this student's experience at least, the word-only facility had the opposite effect to that intended. Rather than enhancing the reading in a positive way, it caused her to dwell on the unknown rather than the known. Initial exploration of the word-only facility (and most students did listen to a few words) might have lead to a decision that listening to individual words was confusing or at least not beneficial.

We do not know, from the design of the current experiment *why* students did what they did (or in this case, did *not* do what they might have done) and we cannot argue conclusively that students would have performed differently on the post-test if they had used the word-only listening. However, the possibility that the listening was perceived as a disadvantage by students (in contrast to our intuition) remains a distinct possibility worthy of further investigation.

5.1.1.8 Post-Test Results Summary

All the above explanations assume that there is some effect of listening while reading on vocabulary acquisition but that due to weaknesses in design, measurement, or interface, these effects did not come to light. The possibility exists that listening while reading does not affect the incidental acquisition (or understanding) of new or low-frequency words (as it was operationally defined) in Japanese learners any more than reading without listening does. It has not been a goal of this thesis to examine reading while listening against a control group who neither read nor listened and read. The literature supports the notion that reading is beneficial for incidental vocabulary acquisition and the intent was to compare the two kinds of reading. However, there are certain findings in this study (the questionnaire data, as discussed in the next section) that suggest that the addition of accompanying sound, and the way in which students had to access that sound, had some negative effects on learners—if not on incidental vocabulary acquisition *per se*, than on students' perception of textual difficulty and textual comprehension. The fact that the post-test means were higher in the control group does not undermine this premise (though the difference was not significant). The questionnaire results to follow in section 5.2 support the notion that the addition of sound is not necessarily beneficial to low level novice users and that this conclusion is independent of the type of interface implemented in this study.

5.1.2 Log Results

The use of the listening facility and the correlation between the amount of use and post-test scores offer an explanation of the nature of the students' reading processes. Hypothesis 2 from Chapter 3 predicted that there would be a significant correlation between post-test results and the amount of word and sentence listening. The correlation between various kinds of listening and test scores was not strong enough to reject the null hypothesis. There was a high degree of sentence listening among all students and across the whole post-test range of scores (on average, students listened to approximately 80% of the possible sentences). On the other hand, the low level of word listening was slightly disappointing. Some students ignored this facility altogether and those who used it did so sporadically. We cannot rule out the possibility that had all the experimental group students listened to 100% of the story and the word listening, they might have performed differently on the post-test.

Despite the fact that the weak relationship between the amount of listening and post-test scores, the log data did show some characteristic patterns of student use. It was not predicted, but careful examination reveals that students can be grouped into at least four different styles. The data can be characterised along these lines (see Appendix 4.1 for examples of the log data).

Table 5.1-1

Categorisations of Experimental Group Listening

Category	1	2	3	4
Label	Beginning to End Once	Once Plus More	Twice	Idiosyncratic
Criteria for categoris-ation	• student listened to more than 50% of the sentences in each page • student read sequentially • student did not go back for a second read or listen	• same as type 1 except students scanned through a second time	• same as type 1 except students read text through twice completely	• student did not read whole text • student did not read sequentially • some students gave up after reading a few pages

These categories will henceforth be referred to by number.

It should be noted that the log data tell us nothing about the quality of the reading and listening, i.e., we do not know what students hear, nor how *well* they were listening, concentrating, or reading. In addition, no similar log record is available for the control group—we only know the number of page turns and the time spent on each page.

5.1.2.1 Background to Navigational Patterning

Hypertext navigational patterning is determined primarily by the navigational possibilities of the specific materials (Horney 1993). In an examination of expert users and their navigational trends, Horney found five different categories in an experimental hypertext, the first and last of which match fairly closely with the patterns found in this study:

- *linear traversal*
  students worked from beginning to end;
- *star*
  students worked then branched, worked then branched;
- *extended star*
  students worked, branched, then branched again;
- *side trip*
  students branched and rejoined main path later;
- *chaotic*
  students' patterns were difficult to categorise.

Patterning in computer-based materials has also been observed by van der Linden (1993). In this study, students were given on-screen grammatical exercises. The computer offered suggestions if mistakes were made, but did not force students to get the correct answer before they moved on. Patterns were categorised as follows (p. 63):

- 'go on till you get it right'
  students attempted to try and retry answers until they came up with the right one;
- 'drill-like'
  students did not try to correct themselves when they got an answer wrong, they just asked for the correct answer and moved on;
- 'once is enough';
  students made one guess then gave up and moved on;
- 'browsing'
  students tried to answer here and there but moved through the exercises quickly and seemingly without purpose.

Again, the patterning is comparable to the patterns found in the present study and in Horney's research—even though the kinds of computer-based material were different. Although the reading exercise in the present research was largely a linear exercise, students did have freedom to listen to the sentences and read pages in any order they preferred. The obvious question is whether certain kinds (or level) of student read with different styles. We cannot say that reading in a certain way facilitates the acquisition of new words (although this is certainly possible).

Horney surmised that his chaotic patterning was an indication of complex goals being sought by experienced users. Van der Linden felt that, given the opportunity to 'cheat', many students would not go through the materials as the designer intended. The 'browsers' and 'once is enough' students in van der Linden's view were examples of students using (what the author considered) less-than-optimal strategies.

5.1.2.2 Implications for the Present Experiment

Obviously, different patterning could manifest itself for any number of reasons. Horney used educational researchers with considerable computer experience. Experienced computer users may jump inexplicably for different reasons than low level language learners. Horney calls the 'random' navigational patterns of his experienced users 'purposeful exploration'. An inexperienced user, conversely, may make 'random' jumps for wholly different reasons (confusion, frustration, etc.). This might be called 'purposeless floundering.' Different levels of student may use different characteristic strategies for different reasons. Although a one-way ANOVA showed no significant difference between any of the categories in terms of post-test scores, a more suitable test for reading ability or comprehension might yield more conclusive results (see Appendix 4.2 for details of the ANOVA procedure). The systematic first reading plus secondary scanning found in category 2 reading (and categories 1 and 3 also appear systematic to a degree), might be characteristic of higher level students when the corresponding computer familiarity is low. It may be worthwhile, in a future investigation, to determine if their is any relationship between pattern-type and reading ability, vocabulary acquisition, computer familiarity, or language level. If different level students use hypermedia materials in characteristic ways, this kind of knowledge would help in the design phase of hypermedia materials (by suggesting what options suit particular students) and in the instructions given to students (by suggesting how they should use the materials).

5.1.2.3 Patterning and Learning Strategies

In Chapter 2, we discussed the implications of patterning in terms of learner strategies and specifically with regard to systematicity. It is reasonably clear from these log data that students approached the experimental task with at least three identifiable strategies. The question remains as to why students chose to read and listen the way they did. Obviously, as was said in Chapter 2, the range of possibilities within the experimental reading was limited. However, the literature seems to show that successful language learners do some things more than unsuccessful learners. Abraham and Vann (1987), for example, found that a successful learner is many times more likely to verbally repeat words or phrases, to inquire, ask for meaning, and take more risks with the language. The unsuccessful learner is more likely to monitor vocabulary use and comprehension while the successful learner just moves on when encountering an unfamiliar word. We would strongly expect, therefore, that the less structured, less repetitious approach characterised by category 4 students indicates the strategy of a less successful learner.

Since several studies have found that excessive monitoring of vocabulary is indicative of poor learners' strategies (e.g., Abraham and Vann 1987, O'Malley and Chamot 1990), one might have expected more individual vocabulary listening among category 4 students. It may be, therefore, that individual word listening was ignored by different students for different reasons. Higher level students, who are more able to understand the instructions, more likely to adapt quickly to new methods, and more likely to use good features repetitively, may have ignored the option because they found it distracting or because they knew (instinctively or otherwise) it would obstruct their comprehension. Lower level students may have ignored the listening because they were not able to understand the researcher's instructions and consequently could not find the feature. If the lower level students had been more aware of the word-listening feature, they might have used it more—possibly to the detriment of their reading comprehension.

5.1.2.4 Pacing and Control

As was discussed in Chapter 2, the speed at which an aural text is encountered is an extremely important correlate to comprehension (Neville and Pugh 1973). To recap, time-compressed text is more difficult to comprehend than time-expanded text; the difficulty of a text can be manipulated without altering its content through changing the speed of its recitation. In Neville and Pugh's research the pace was controlled for the duration of the text in an unbroken manner. In Chapter 2 it was suggested that computer-based speech would be easy to comprehend if students could control when they listened. However, it is apparent from these data that the positive effect gained from pacing (slowing) an entire text, with no user-control, is not immediately found in a situation similar to the present one.

There are several possible reasons for this. First, the text within each sentence occurred at a natural speed. One might have expected that the option to listen repetitively to a sentence would make it more comprehensible—perhaps even overcoming the (apparently) fast pace of the experimental reading. However, the results of this experiment, and those of Neville and Pugh's experiment, seem to indicate that the pace of the recording is more critical for vocabulary comprehension than the option to listen repetitively. One other factor to consider is that the pacing found in Neville and Pugh's experiment was continuous. It forced the reader/listener to read from beginning to end and as a consequence, students were obliged to use a strategy characteristic of successful readers: they had to skip unknown words without undo monitoring and do the best they could with what they heard. There is much to be said for permitting students to learn in their own way. It may not be a bad idea, however, to 'push' students to read in a specific way—especially if one can vary the speed (and thus the difficulty) at which text is presented. It is possible that the present experiment failed to produce the expected result because it permitted poor readers to read with their unsuccessful strategies and/or because it encouraged good students to use 'stop and consider' strategies of less successful readers.

Improvements could be made to this experimental design by requiring the group to listen to the whole text without interruption. The text could be expanded or compressed by adding space between or within words and measurements in vocabulary acquisition could be taken to test the importance of reading and listening without interruption. Doing a future experiment in this way would clarify two uncertainties: (1) the question of pace and vocabulary acquisition; and (2) the question of user freedom versus controlled pacing. With regard to the first question, it would be beneficial (for the materials designer and the teacher) to know if speed affects vocabulary test scores over a short-term experiment or over repeated exposure to words in various context. If the results of such an experiments suggested that pace significantly affects vocabulary acquisition, it would be interesting to build in a 'speed control' into on-line reading materials. Students would then be able to control the speed themselves, or have their pace controlled by a teacher.

With regard to the second question, it would be beneficial to know, with specific reference to reading and listening, to what degree freedom affects post-test results. Should students, for example, be permitted to choose between continuous reading and sentence-by-sentence reading or should they be required to read in a certain way at a certain pace? It would be interesting, therefore, to design an experiment where some readers were forced to read one way, others in another way, and still others were permitted to choose between methods. The post-tests could include affective factors such as enjoyment or anxiety as well as tests for vocabulary or comprehension. This question of user/computer control is at the heart of hypermedia research and assumption (see Figure 1.4-1). It would be useful to know how students could be forced to use beneficial strategies and whether or not the subsequent loss of freedom is worth any gains in learning.

5.1.3 Context and Guessability

The likelihood of students determining lexical meaning from context depends as much on the richness of the context as the students' ability to infer. In order to find out how rich the context was and how much students could be expected to learn from it, the story was given to a group of Japanese students who did not participate in the computer-based reading and listening experiment. The post-test target words were removed from the text and replaced with gaps. Students were asked to write (in Japanese, English or both) the meaning of the word that was removed. In order to avoid removing too much context, only one third of the target words were removed from three versions of the story. The researcher attempted to leave two or three sentences between gapped words. Each version was then given to a different group of students.

Obviously, the results from these kinds of data only reveal what the context tells us about the word. Students in the main experiment had the word in front of them and would therefore have more information than students in the post hoc experiment. The computer-based readers would, for example, have information about the words' parts of speech and morphemes within the words. Students might have had previous encounters with the words or they might have known related words with similar parts. Nevertheless, a cloze test such as this should be a good indicator as to how much the context could be expected to contribute to students' comprehension of target words.

The texts, versions 1, 2, and 3 (N = 17, 15, 22 respectively) were given to 54 first and second year Japanese university students. Students were given written instructions and up to 40 minutes to complete the rading and gap-filling. All students finished the exercise within the time limit—although not all students filled every gap. The results (see Appendix 6.2) showed that 60% (29 out of 48) of the word meanings were guessed by one or more students. Only on two occasions were the actual target words guessed, but this is not surprising considering the words were chosen on the basis that they were probably unknown to Japanese university students. It should be noted that in some cases, the author's experience with Japanese students' vocabulary was used to determine their intended meanings. The word 'green' for example, is usually used by Japanese students to mean 'greenery' or 'shrubbery' when used as a noun (see Appendix 6.2: *tamarind*).

Very few word meanings were guessable (only 3) by a majority of students reading each version. This suggests that even though the context provided enough information to guess the words, for the average member of the population under examination, most word meanings were not inferable from the context. The possible implications from this are twofold: (1) the students' general English level may not have been high enough to comprehend the context or (2) this population might not have been skilled with the strategy of infering from context. Of the two choices, the second seems most likely to this researcher. In many classroom situations, Japanese students have proved to be reluctant guessers. Over the past year, most of the researcher's classes have had to be taught to use context. Intuitively, the researcher has observed that successful guessing often takes a good deal of practice for Japanese university students.

The data from the main and the post-hoc experiment suggest that one can not generally expect (with Japanese students) that reading and listening will provide gains over the short term in lexical understanding through context. However, any future investigation of this sort should include pacing as an independent variable and should also include a lengthy period of training for context-based guessing. A future implementation of this experiment might also include a design where all students are in both the control and experimental groups. Post-tests could be administered after both the control and experimental readings and comparisons could be made at both stages.

5.2 Questionnaire Results

5.2.1 Preliminary Discussion

5.2.1.1 Data Reduction

The categories in item 9 were collapsed for two reasons: (1) to make a χ^2 test possible and (2) to concentrate on either the agreement or disagreement with each questionnaire item. By collapsing the data, discrimination between responses is reduced. Nevertheless, depending on the statement, agreement or disagreement can still be a meaningful distinction and we believe that the integrity of the questionnaire data is not lost by collapsing it to 1, 2, or 3 categories.

The reduction of categories in items 19 and 20 requires slightly more discussion. Again, the manner in which students responded in each item's category do not permit the use of a χ^2 test across all categories. However, the division between more or less than 50% comprehension is arguably imprecise. The discussion of items 19 and 20 should therefore be understood in this light.

5.2.2 Difficulty (plot and vocabulary)

The data from items 5 and 6 suggests that the addition of sound to the reading exercise did not make the words or plot easier for the experimental group. Indeed both groups considered the plot and vocabulary to be difficult.

We did not predict that there would be so many novice users in the questionnaire sample. Indeed it was expected that most students would have at least passing familiarity with a mouse and that the training period would be simply a re-familiarisation and a means of initiating students to the specific way in which the mouse would be needed (pointing at the beginning of sentences and looking for linked words). It was optimistic to expect complete beginners to master the use of the mouse in the short training period provided.

The correlation between experience and difficulty in the experimental group supports what was suspected after analysis of the post-test sample—that in novice users the addition of sound produces an adverse effect on the learner in terms of their perception of context/lexical difficulty of the reading text.

5.2.3 Comprehension (plot and vocabulary)

The idea that some students react adversely to sound is supported by the data from item 19 (plot comprehension). The preconception of the researcher was that the inclusion of sound would produce higher ratings of comprehension than plain reading. However, the control group indicated a higher level of plot comprehension than did the experimental group. The implications from these data are twofold:

1. Students who use a talking book may initially feel less secure about their level of comprehension than those who simply read. The data do not show a negative attitude towards this exercise and most students agreed that they would want to use the computer for this kind of thing again. However, the lack of confidence students feel about how much they understand during an activity may still reduce the likelihood of them using those materials frequently. This is especially pertinent in a self-access situation (where this program is most likely to be implemented) because affective factors such as these help to determine students' choice to use the materials.
2. If the results indicating the level of perceived comprehension is matched by a similar level of actual comprehension (as might be measured by a comprehension test) then it would be counter-productive to use a talking book (if an enhancement of comprehension is the goal of the inclusion of sound). The generalisability and power of this conclusion must of course be evaluated with the present sample in mind, i.e., low-level Japanese university students mainly inexperienced with computers. Consideration should also be given to the limitations of imprecise data as suggested in section 5.2.1.1. Moreover, since the

questionnaire was short and there were only two items pertaining to comprehension, the conclusion would have to be supported by future investigation.

5.2.4 Enjoyment

It was immediately obvious from informal observations that those involved in listening were concentrating well. On several occasions, other teachers who happened to be in the same room while the experiment took place, noted how absorbed the experimental group students were. Several other times, when the experimental group students were interrupted to be told to stop, they were visibly surprised, as if startled from deep concentration. Questionnaire results showed that the experimental group on the whole, enjoyed the reading activity more. As other researchers have noted, computer-based materials are prone to producing a 'novelty effect' (e.g., Krendl and Brohier 1992). We would expect, because of the sample's relative inexperience, that a 'novelty effect' affected the results to some degree. However, the strength of the response to this question cannot be accounted for fully by the newness of the exercise. One would expect, for example, that if there was a novelty effect influencing the item 9 results, then there might be a corresponding effect positively influencing perceived comprehension (items 19 and 20 on the questionnaire). In fact, it was the control group who reported significantly greater plot comprehension (with the 50% division) than the experimental group. The enjoyment of the activity does not necessarily correspond to whether or not students felt they understood more than 50% of the story's plot. In fact using all response categories, the Pearson correlation showed no significant correlation between items 9 and 19, suggesting that students were not judging their enjoyment by how much they comprehended the text. In other words, it is possible for students to enjoy this kind of reading activity without feeling that they understood it well. If it is true, as Blattner (1993: 77) argues, that "audio is one of the most powerful methods of engaging the mind..." and that there is nothing "more enticing than a computer that speaks to you" (Blattner 1993: 78), then it may be possible to also claim that the addition of the audio context made the reading more enjoyable for the experimental group. It may also be true, therefore, that reports of enjoyment are not entirely due to a novelty effect.

There is some other evidence suggesting that a novelty effect was not responsible for increased enjoyment in the experimental group. One would expect that if a strong novelty effect had been present, there would have been much stronger disagreement with item 10 (Hypothesis 7). In fact, as has been said, there was very little disagreement with the notion that a normal book is preferable to the computer. It appears, therefore, that if students in the experimental group enjoyed the reading because it was new, this tendency was not evident in other related items where it might be expected to emerge.

Having said this, the notion of a positive novelty effect on item 9 results should not be discarded entirely. Open-ended responses from the control group produced some positive comments about the reading exercise. Three students (SN: 35C, 2C, and 3C) noted that they would not change the exercise at all and that they enjoyed it immensely. It is difficult to imagine even the most enthusiastic proponent of hypermedia thinking that reading a story, consisting of nothing more than text on a computer screen, was 'fine as it is.' Interestingly, not a single control group student commented that they could have just as easily read the story in a book (which seems obvious to the researcher) and only one student (SN: 32E) noted that reading in books is easier than reading on a computer. There apparently was either some novelty effect in the control group or several enthusiastic technophiles in their number.

Students did enjoy reading with sound—despite the fact that many found it difficult. Student responses in the open-ended questionnaire from the experimental group included the following observations:

- they enjoyed it;
- they believed the exercise was interesting;
- they would like to try this kind of thing again;
- they believed the exercise to be effective;
- the computer makes reading exercises more interesting;
- it was extremely helpful to be able to listen repeatedly;
- it was helpful;
- it is a good way of improving listening skills;
- it was as natural as reading a book;
- using the computer helps students to concentrate more;
- it is good for self-study.

There were several negative comments—mostly to do with the design of the materials themselves—which will be discussed in Section 5.3.

5.2.5 The Preferred Medium

Despite the comment by one student in the control group that he preferred the computer to a book as a reading medium (SN:12C), a significant number of students in the combined group did not disagree that traditional reading materials are preferable. Since students on the whole, enjoyed using the computer, this seems somewhat of a contradiction. However, most computer users would agree that reading on a computer is tiring for the eyes and not as pleasant as reading from paper. The screen flickers, and the print appears at a much lower resolution than the printed page. Furthermore, most people read looking down at their page, and probably sitting comfortably. Computer users look across a desk, sitting in an office chair. The open-ended sections of the questionnaire confirm that the present sample also felt that reading on screen was tiring for their eyes. Five students felt strongly enough about it to comment on this point (SN: 39E, 17C, 35C, 2C, and 9C). It is therefore not unreasonable to expect some resistance to computers as a reading medium. A high resolution screen would be an improvement. However, even the best screens fall short of the quality of the printed page.

5.2.6 Predispositions

It is understood that any correlation between students' predispositions and enjoyment of the experimental exercise would not be causal (i.e., a significant correlation does not imply that enjoyment of traditional reading *caused* the student to enjoy the computer-based activity), but we wanted to determine if there could be a high level of computer-based reading enjoyment without a corresponding high level of traditional reading enjoyment or liking of computers.

Items 10 to 14 addressed the questions of students' predispositions. There was found to be no significant correlation between the enjoyment of the computer-based reading and: (1) students' enjoyment of non-computer-based reading and listening (i.e., tape recorder and teacher-based reading and listening) or (2) enjoyment of reading in Japanese. Furthermore, there seemed to be no relationship between students' perception of the difficulty of reading in English and their ratings of vocabulary and plot difficulty. From these data at least, it appears that students in this population could enjoy the activity without liking reading in Japanese and without enjoying reading while listening with traditional methods. In terms of difficulty ratings, students' impressions that reading in English is difficult are not necessarily linked to their impression that the activity itself is difficult. It is important to remember, however, that students in both groups thought the plot and vocabulary were difficult. In no way can it be inferred from these data that students who think reading in English is difficult thought that reading with the computer was easy. On the contrary, the opposite may well be true—students who think English reading is not difficult may have thought this particular mode of reading was difficult.

However, there was a strong correlation between students' liking computers and their enjoyment of the story. Generally speaking, the more a student liked computers, the more likely she was to enjoy the experimental reading.

The link between liking computers and enjoying the story is borne out in particular by one student's (SN: 35C) response in the open-ended data. This student noted that,

> I do not think the program should be changed in any way. Adding things (that are unnecessary) would make the screen too decorative (and possibly bad for the eyes). I found the lesson really interesting and I guess it is mainly because I am interested

> in computers. Reading books can be boring—sometimes it makes me really sleepy, but the same thing can be made so enjoyable with the computer.

This student's enthusiasm perhaps explains in part the trend suggested by the strong correlation. It is possible that students who like computers would respond positively to almost any computer-based activity (the fact that the above student was in the control group strengthens this suspicion). However, it is also possible that students without the same enthusiasm enjoyed the activity for its own sake. Several novice users commented that although it was the first time they had used a computer, it was enjoyable and worthwhile. For example,

> This is my first experience with computer-supported reading in English. It was as natural as reading a book. I also like using the mouse. (SN 22E)

> It was interesting because I used a computer for the first time. (SN: 24E)

> It was new to me and I enjoyed it very much. I hope we'll have more opportunities like this. I believe the program is a really effective way to improve my listening skills....(SN: 26E)

So the fact that students who like computers tended to like the exercise does not rule out the possibility that students who do not particularly like computers also enjoyed it and would enjoy more activities like it.

5.3 Materials Design Implications

5.3.1 Preliminary Discussion

So far, we have rationalised the present design of the mouse pointing and screen layout primarily on the basis of its experimental reliability. It is possible that experimental concerns have caused the usability of the materials to suffer. It has been a goal of this study, however, to produce materials that would have some use in a normal self-access centre. It was for this reason that an authentic reading text was chosen and that the 'talking book' format was used. The log data and the open-ended questions suggest that several changes could be made to improve the design and usability of the reading materials and that these changes might be developed into more general guidelines for the development of this type of materials. This section will discuss the program as a teaching tool. As the changes are discussed, the experimental concerns will be borne in mind but the primary concern will be improvements for the sake of pedagogic functionality.

Quentin-Baxter and Dewhurst (1992: 181) argue that the efficiency of a hypermedia environment

> ...may be regarded as the effectiveness of the user interface and functional design to provide a student with the best opportunity to obtain all the relevant subject information that the program is able to offer.

The log data demonstrate that users did not 'explore' the word-only listening facility. The very low rate of use of this facility might be explained in three ways: (1) students might have chosen not to use the word-listening (as discussed above); (2) students may not have been aware of the word-listening; and (3) students may have not had enough training time. Numbers 2 and 3 will be discussed in this section.

As was said in chapter 3, we attempted to make the reading as book-like as possible. It was also felt that making the words italic, bold, or underlined (which is the usual way of indicating a linked word to the reader) might change the natural reading strategies of the user—causing him or her to pause or mentally emphasise the words more than they naturally would (and it is possible that the exercise had this effect anyway). In the end, the only indication that words were linked was that the on-screen cursor changed from a vertical line to an arrow when passing over the linked words. Generally speaking, from an experimental perspective, the decision to keep the linked words unmarked was warranted. In retrospect, however, from a pedagogic perspective it appears that the changing arrow was not enough to remind students that the word-only option existed. In a self-access situation, this would be unacceptable.

There are many possibilities for design improvements in this area. For example, it would be possible to program the exercise in such a way that highlighting of the linked words could be turned on or off. This would facilitate a first reading of the text in a natural and unbroken manner and a second or cursory reading with the linked words appearing in a different colour/bold/etc.. If the students were 'forced' to read through once and could only get the multimedia elements on a second pass, it would encourage a more thorough reading of the text and increase the possibilities of incidental vocabulary acquisition. Linked words might also appear in the margin, as the text is heard, a button might appear in the corner of the screen, or a combination box might appear showing an alphabetical or context-sensitive list of linked words. Where experimental reliability is not a concern, other hypermedia features would most certainly be added (see table 5.3-1 in section 5.3.3). There are many possibilities.

Students also complained of poor response times with the mouse and that it was difficult to use (students 28E, 20E and students 30E, 32E, 43E respectively). The response times are usually a function of the processor speed and are thus not within the designer's control. However, the mouse use could have been simplified had the size of the 'hotspot' been larger. Students were observed to spend some time trying to find the linked first letter of the sentence. The problem usually solved itself as students realised that the cursor changed when the link was active but there was certainly some initial frustration trying to get the sound from each sentence. Again, the experimental design called for an unbroken screen but a better way of eliciting sound would be to have a stop/start panel—resembling a tape-recorder's buttons—on screen. The sound could be matched to some sort of visual clue indicating the position in the text and the difficulty of pointing, finding, and clicking would be eliminated. It should be noted that even if we had chosen this kind of visual device, it would have been prohibitively difficult with the software and expertise available at the time.

5.3.2 Implications

The experimental constraints are possibly at odds here with any pedagogic improvements we may wish to implement in a future system of this nature. As we reduce the number of variables so that the experiment becomes more reliable, we reduce the likelihood of the materials in question being usable in a pedagogic context. As it is, the current exercise is not without functionality; many other 'talking books' make use of similar interfaces. However, experience with other multi/hypermedia programs has shown the researcher that users need all the facilities to be conspicuous or they simply will not use them. In the computer field, the most extreme examples are the system interface used by DOS (a command driven interface) and that employed by something graphical like Windows (a graphical user interface). Most of the file features of Windows are available in DOS, but Windows makes file maintenance easy because all the options are visibly available—without learning a command to make them work. The alternative to the 'make it obvious' approach is time, i.e., training and long-term experience (Hazari and Reaves 1994). The present exercise was not as difficult as a command-driven interface. Nevertheless, its manner of operation may not have been obvious enough for users to take full advantage of it.

The choice of interface complicates the evaluation of multimedia materials. The difficulties of the CDI and the superiority of the GUI (in terms of learning time) are now reasonably established in the literature (see Hazari and Reaves 1994, for a thorough comparison of the two types of computing). Hazari and Reaves write that (p.228),

> Computer anxiety, prevalent when microcomputers were first introduced, is now giving way to a user-friendly environment for students to use computers as a tool to deliver their message without spending much time and effort learning about the medium.

However, the differences among hypermedia interfaces are more subtle than the differences between GUI and CDI. Therefore, the evaluation of the way in which users interact with hypermedia materials is much more difficult. The fact remains that there is still a lengthy training period required to become skilled with graphical interfaces—easy as they are in comparison to past systems. In particular, manipulation of the mouse is a difficult and sometimes time-consuming skill to acquire—although most users do become skilled eventually.

5.3.3 Student Expectations of Computers in Language Learning

Interestingly, and despite their computer inexperience, these students were aware of the possibilities within multimedia (incidentally, in this researcher's experience, students are much more aware than their teachers). Student responses to the three open-ended questionnaire items ranged from enthusiasm (SN: 35C) to reserved disapproval (SN: 36E). These results seem to indicate that the students had certain expectations about what the computer should be able to do in general and what it should do for them in a language learning context. Table 5.3-1 summarises the requests made by students:

Table 5.3-1

Summary of Student Requests for Additional Features

Request	N of students from each group who made request. (C = control; E = experimental)	N of students to make request
On-line dictionary or equivalent	4C, 6E	10
graphics or pictures	5C, 4E	9
games	3C, 2E	5
Varied topic	4E	4
Varied difficulty level	3E	3
animation	1C, 2E	3
quizzes	3C	3
wanted to hear a teacher's voice at the same time as reading	1C	1
wanted to hear a tape at the same time as reading	1C	1

In other words, these students expected more than just a one-dimensional activity (as the reading-only could be described) or even a two dimensional activity (like reading and listening). Students expect hypermedia. Unfortunately, the experimental design in this study did not permit the use of *true* hypermedia. It was, in fact 'stripped' hypermedia. However, the data suggest that the inclusion of sound, while not necessarily beneficial in terms of strict measurements of vocabulary gains or perceived comprehension, adds enjoyment to short stories. This should be encouraging for proponents of audio in hypermedia. The benefit of added enjoyment is straightforward: a student who enjoys an activity will do it again; the more reading a student does, the more vocabulary they will encounter and eventually acquire.

5.3.4 Interaction/Interactivity

This stripped hypermedia can also be discussed in relation to interactivity and interaction. The interactivity (i.e., the amount of shared control between user and computer) and interaction (the type of communication between user and computer) of the present materials can be evaluated on the basis of the definition established in section 1.4.3.3 and figure 1.4-2, respectively. In terms of interactivity, the students in this experiment were not directed in any way by the computer. They were limited by the number of available features and the way in which they could use them, but they were free to go through the text in any way they chose. Since the user had complete control of movement through the document, the interactivity, as defined by figure 1.4-1, was low. Similarly, the interaction was one-sided—it consisted entirely of *enhancement* (see figure 1.4-2), ignoring the other desirable features of *correction* and *suggestion*. In terms of pedagogical validity, both these facts should be seen as limitations of the present materials if they were to be used in a classroom—assuming that high interactivity and variety of interaction are more effective and desirable pedagogically.

5.4 Summary

This chapter has been concerned with both vocabulary acquisition and hypermedia design. From the perspective of vocabulary acquisition, the data show us two things: (1) incidental acquisition of new words through listening and reading probably occurs in very small increments; and (2) if those increments are measurable over a short time period, the experimental design must carefully control the variables of previous computer experience and student level. Intuitively, one would expect that the addition of sound would make *comprehension* of the story easier, as measured by our comprehension test. The data, however, suggest that this is not the case. There is (as suggested in the post-test sample) either no effect on short term vocabulary acquisition or (as suggested in the questionnaire sample) a negative effect on self-reports of plot comprehension. The log data show that students adopt a variety of styles when using a talking book. The possibility exists that there is some relationship between the type of style a reader uses and his/her language level, and tentatively, that the approach a student takes to the reading will affect their ability to learn new words.

From the perspective of hypermedia design, this chapter has shown that students have somewhat higher expectations of computer assisted language learning than the present reading exercise offered. Many complained about the way in which they were required to elicit sound (i.e., the interface). At the same time, many students wrote that they thought the computer-based reading was useful. Improvements could be made to the interface but the data from this thesis suggest that there is a great deal of frustration involved in the learning of any graphical interface — even though it is markedly easier than learning commands. This finding notwithstanding, users in the experimental group reported a much higher level of enjoyment than did the control group. This suggests that sound may be a legitimate means of encouraging reading, assuming that greater enjoyment would lead to greater use.

Table of Contents: Chapter 6

CHAPTER 6
SUMMARY AND CONCLUSION

The first section of Chapter 6 will summarise the literature review, experimental method, results, and discussion. The second section of this chapter will discuss some of the conclusions drawn in Chapter 5, and the last section will make suggestions for future research.

6.1 Summary

In Chapter 1, we discussed the history of computer assisted language learning in terms of its roots in computer assisted instruction and educational technology. It was argued that within CALL, hypermedia could be used as a means of creating stimulating and useful materials, but that to date, there was little research-based evidence suggesting the effects of specific implementations of hypermedia on specific pedagogical interests. It was proposed that a good starting point would be the investigation of computer-based reading and listening and its effects on vocabulary acquisition, comprehension, and reading enjoyment. Chapter 2 discussed the history of vocabulary in language teaching, gradually narrowing the focus of the discussion to incidental vocabulary acquisition through reading and listening. It was discussed how previous studies had shown some measurable acquisition of new or unfamiliar words after reading a text that contained the words in several contexts.

We then posited that the addition of the spoken text might produce measurable changes in vocabulary acquisition where new or unfamiliar words appeared only once or twice in an authentic reading context.

In Chapter 3, these suppositions were refined into null hypotheses and alternative hypotheses and an experimental method was developed to test them. The experiment consisted of two parts. In the first part, a control and experimental group were asked to read a short story. The groups differed only in that the experimental group had the option of listening simultaneously to the text. Both groups were given a multiple choice post-test with the intention of measuring the difference between group means. The computer logged the amount of listening that students did. It was hypothesised that there would be a correlation between the amount of use and post-test scores.

In the second part of the experiment, a different group of students were divided between control and treatment groups, presented with the same text as the post-test students, and then asked to complete a questionnaire after the reading. The questionnaire examined students' attitudes towards the procedure (including enjoyment, reading medium preference, and liking of computers) as well as examining students' self assessment of vocabulary and plot comprehension, ratings of vocabulary and plot difficulty, and self ratings of previous computer experience. It was hypothesised that there would be significant differences between groups in their various responses to these questionnaire items and the null hypothesis that there would be no difference between groups was tested.

The experimental procedure discussed in Chapter 3 was conducted using a sample of Japanese post-secondary English students who had come to the University of Edinburgh for three-week summer courses. Fifty-nine students participated in the post-test experiment; an additional 43 students were recruited for the questionnaire part of the design.

Chapter 4 showed the results of the various statistical procedures used to test the null hypotheses developed in Chapter 3. There was found to be no significant statistical difference between control and treatment group means on the post-test and only a weak correlation was discovered between the amount of listening and post-test scores. However, data from the questionnaire sample suggested that there was a significant difference between groups in their ratings of enjoyment of the program—the experimental group claimed to enjoy it more. However, tests conducted on students' self ratings of comprehension suggested that considerably more control group students reckoned they understood more than half of the story's content than did experimental group students. Other analyses showed no difference between groups on plot and vocabulary difficulty (the combined group did not think that the plot and vocabulary were easy), no difference between groups on vocabulary comprehension, and no difference between groups on reading medium preference (the combined group did not prefer the computer to a normal book). Furthermore, there was no difference between groups' desire to use the computer again for a similar task (the combined group responded that they would want to), no difference between groups for their opinion of the computer as a learning tool (the combined group thought it was a good tool), and no relationship between previous experience and enjoyment.

Other hypotheses examined the possibility of relationships between the following: a predisposed liking of computers and enjoyment of the experiment; a predisposed liking of reading and enjoyment of the experiment; a predisposition to like reading while listening and enjoyment of the experiment for the experimental group; and a difficulty with reading and difficulty with vocabulary and content in the experimental reading. Of these possibilities, one null hypothesis was rejected: that there would be no relationship between liking computers and liking the experiment. The two items dealing with students' previous computer experience showed that, on the whole, the combined group was inexperienced—most students were either complete novices or infrequent computer users. Data from the questionnaire also included open-ended responses to three items. Students were quite forthcoming in their responses to these items (they answered in Japanese and they were subsequently translated for the researcher) and the data provided a good deal of insight for the discussion in Chapter 5.

Chapter 5 discussed these results from four perspectives: (1) why there was no measurable effect in the post-test sample, (2) how inexperience might have affected the results, (3) what the implications of between-group differences in enjoyment might be, and (4) how the materials could be improved to make them suitable for classroom use. It was argued that the post-test experimental design was weakened by the one-off nature of the procedure. It was also noted that if incidental vocabulary acquisition does occur during reading while listening, the incremental increase is not sufficiently greater than 'plain reading' incidental acquisition for it to be measured with a multiple choice post-test. Had the text been longer, we argued, and had the words appeared in a variety of contexts, there may have been a more marked difference between groups. In any case, from these data, it is not possible to say that there will be any difference between similar Japanese groups who read and listen or simply read in incidental vocabulary acquisition (as it was operationally defined).

The second concern of Chapter 5 was that students' previous computer experience might have adversely changed the effectiveness of the reading exercise. The questionnaire data did not demonstrate any clear link between previous experience and enjoyment. However, the possibility remains that in this experiment there was a link between previous experience and general feelings of frustration—as there has been in many other studies. For example, qualitative observation of the students while they undertook the experiment, and the open-ended responses from the questionnaire suggest that there was a great deal of frustration with the interface and that more experience might have improved students' ability to take advantage of the system.

With the questionnaire it was argued that the significant difference between groups on the question of enjoyment might have been due to a novelty effect but that it was just as possible that the addition of sound was appreciated by students for its own sake. It is hoped that further research in this area might support this finding and that we might find further evidence that hypermedia features motivate students to read more.

The open-ended section of the questionnaire provided most of the data for the last section of Chapter 5, which was concerned with suggestions that students made for improvements to the reading. We have contended from the beginning of this thesis that the experimental materials should be usable without extensive modification in a self-access centre. Many students commented that reading while listening was beneficial and good for listening practice, reading practice, and that it was enjoyable. However, the data also suggest that the materials could be used as they are but that the addition of various other hypermedia features would make the materials more robust (i.e., attractive, usable, and helpful). The open-ended responses indicated that students expect a great deal from modern computer-based materials. Many students commented that on-line dictionaries, games, video, and music would be greatly appreciated. We conclude, therefore, that the idea of 'stripped' hypermedia may be useful for experimental purposes where variables must be controlled. In a real self-access centre, however, the materials could be improved with the addition of features like those suggested by students. For example, in Chapter 2 it was suggested that on-line glossaries can be used in certain contexts as an efficient means of improving reading comprehension (cf. section 2.4.2 and Leffa 1993). One could easily envision materials that included both on-line glossaries and spoken text. One could argue (perhaps simplistically) that glossaries would enhance comprehension and the sound would enhance enjoyment. Further research is of course needed to examine the effects of combined hypermedia elements.

6.2 Conclusions

In the introduction to this thesis, we stated three aims for the study. First, we argued that generally speaking the research should suggest one or two concrete ways English can better be taught and learned. Second, it should use the insight gained from theoretical research to improve computer-based materials, and third, it should inspire further investigation and development in the use of computers for language teaching. With regard to the first goal, we may conclude that the addition of sound, while not producing an obvious effect on vocabulary acquisition as it was operationally defined, probably increases student enjoyment of reading. Some qualifications are worth noting, however. First, the definition of enjoyment was straightforward and simple. Students were asked if they enjoyed the materials and no other measurement was taken. Second, inexperienced computer users may experience frustration with computer-based materials and this frustration may reduce the positive effects of the materials. Despite the comparative simplicity of graphical interfaces, there is probably no way of allowing novice users easy interaction with the computer. In this research, the difficulty that new users would have with the mouse was underestimated. Another unexplained finding in this research is that students who read without sound reported greater content comprehension than students who read and listened. Considering that our initial intuition was that listening would improve self-reports of comprehension, this result is somewhat confusing. It could be that the additional complexity of the reading and listening interface produced some consternation in the students. The fact that the interface might have been difficult to understand may have lead students to report that the text itself was difficult to comprehend. On the other hand, the way in which the data was collapsed in the measurement of comprehension may have lead to anomalous results. This finding, while interesting, should be investigated in a future study that addresses the issue of comprehension more directly. Such a study, in addition to students' own estimation of their understanding, should include questions, either multiple choice or open-ended, that objectively measure comprehension of the text.

With regard to the second goal, there are many implications for improvements to computer-based materials. The first implication is an extension of the problem of previous experience mentioned above. The interface of any computer-based materials should be simple and the available features as obvious and accessible as possible. Inexperienced students will probably not 'discover' hidden features, even if they are given explicit instructions at the outset of that activity. The fact that sentence-listening (which was obvious and easy to find) was used extensively, and the word-listening (which required some investigation to be found) was hardly used at all, supports this notion. The second implication for materials design is that students have high expectations. Japanese students, who may not have actually seen a language learning program before, have an idea of what should be possible and what they would like to be able to use. However, the coordination of all these features in an organised and user-friendly fashion is another matter altogether and could be the topic of a future study. Visual interfaces, for example, need to be compared; experiments with type-face, colour, and screen layout need to be examined; and physical interface devices need to be evaluated on the basis of simplicity of use.

The final goal, that of inspiring future interest in computer-based language teaching, is one that can not be immediately assessed. However, this may be an appropriate point to offer some initial suggestions.

6.3 Future Research

There are many questions raised from this research that might be addressed here. First, it is still not known if reading while listening is conducive to incidental vocabulary acquisition. A more complex design, where students are exposed to words in many contexts and on several occasions might indeed show that reading while listening is a good source of second language vocabulary. Second, a different kind of testing procedure might reveal that the effects, though significant, can be measured only in small increments. More differences might be observed if the question of comprehension were addressed more directly (as mentioned previously) rather than through the learner's own estimation of his or her understanding. The first step of future research should therefore attempt to assess the question of incidental vocabulary acquisition with different experimental procedures and measurement tools.

From a more general perspective, one could argue that the difficulty in the current state of theory and research is that there seems to be a dichotomy between advocates of incidental or passive acquisition and proponents of learning through teaching or instruction. It has not been the purpose of this thesis to show the superiority of either of these approaches over the other. On the contrary, it is felt that while learners probably do acquire vocabulary incidentally over time, it is also likely that they learn from being taught or from taking part in their own learning strategies (be it some form of memorisation, association, or even drill). Future research should therefore concentrate on methods that take account of various learner strategies.

Ideally, a kind of 'hybrid' approach to incidental acquisition is envisioned for an optimal effect. Recently, the researcher has begun to investigate the notion proposed by researchers such as Hulstijn (1992) that guessing words from a context is enhanced by giving students target words and then adding glosses in the margin of the text. Hulstijn found that among several possibilities on paper (including simple definitions, synonyms, or nothing at all) the best way to introduce new words is to provide multiple choice alternatives in the margin so that students need to use more 'mental effort' to choose the correct answer from the context. In using more mental effort, they would concentrate more on the context of the text and be more likely to retain the meaning of the word over time. The problem with paper versions of exercises that make use of this method is that wrong answers may also be arrived at by using the context. If readers learn through overt contemplation of the context, but make the wrong choice, then any benefits to the retention of the new word are applied to the learning of the wrong meaning and students are left to then 'unlearn' this meaning at some time in the future.

However, it is this kind of activity where the computer is in its element—and where no traditional materials can match it—on two levels. First, the computer could provide immediate feedback through either a hint, explanation, or by simply telling the reader they are wrong. The computer is well suited for providing feedback on vocabulary because there are a limited number of correct answers (though there may still be quite a large number of possibilities). Second, the computer could provide a range of theory-supported methods to enhance the acquisition of words through reading. With what is known from the present research about enjoyment, the addition of sound might further enhance the effects of on-line glosses or marginal lexical choices (in keeping with Hulstijn's mental effort hypotheses). These kinds of method could be fairly reliably researched—with a multiple choice post-test in a one-off design, or with some other method like self-reports of word knowledge or Meara's (1992) matrix model of vocabulary acquisition and loss.

Chapter 1 presented several issues that are in need of research in CALL. Among these, issues of interface design, enjoyment, comprehension, and vocabulary acquisition have been addressed to some degree in this thesis. However, a central issue in the thesis is that the experimental model is a 'stripped' hypermedia design and that research should progress with the gradual addition of features so that an accurate understanding of each hypermedia element can be understood first on its own and later in interaction with other features. Eventually, research might be done on full, hypermedia materials and evaluation made (perhaps through log files) of various user activities. Their amount and kind of use could be assessed against quantitative or qualitative criteria like post-tests, questionnaires, think-aloud protocols, or interviews.

In discussions with other hypermedia materials designers, two queries have repeatedly surfaced with regard to computer-based materials. In the simplest terms, the questions are: (1) do the materials work, and (2) how can I make them work better? This thesis exemplifies the author's predisposition for straightforward answers to pedagogical questions. This thesis answered *Do they work?* by framing the practical question, *do the materials enhance vocabulary acquisition?*; and *How can I make them work better?* was contextualised in the question, *what do students think of these materials?* If one was hoping for a strong effect on Japanese students' short-term vocabulary acquisition, then the materials (or the measurement tools) were inadequate to produce or determine such an effect and the results were disappointing. If, however, one was hoping for materials that students would enjoy and find useful, then the results suggest that computer-based simultaneous reading while listening fits these criteria.

When language materials are the subject of research, one is never looking purely for quantifiable effects; nor is the appeal of the materials ever the sole concern. It is possible to have stale but effective methods and also to have very enjoyable but useless materials. Not many teachers would desire either of theses choices. This research directs us to consider computer-based reading while listening from these two perspectives. On the one hand, we can consider it an appealing tool for language teaching. On the other hand, we are still left with many intriguing questions about the nature of vocabulary acquisition and the potential of computer-based reading while listening.

BIBLIOGRAPHY

Abraham, R.A. and R.J. Vann. 1987. 'Strategies of two language learners: a case study'. In Wenden and Rubin (eds.). 85-102.

Adams, M.J. and A.W.F. Huggins. 1985. 'The growth of children's sight vocabulary: a quick test with theoretical considerations'. *Reading Research Quarterly*. 20/3. 262-281.

Aitchison, J. 1987. *Words in the mind.* Oxford: Basil Blackwell.

Alatis, J.E., H.B. Altman and P.M. Alatis. 1981. *The second language classroom: directions for the 1980's*. New York: Oxford University Press.

Alderson, J.C. 1984. 'Reading in a foreign language: a reading problem or a language problem?' In J.C. Alderson and A.H. Urquhart (eds.). 1-24.

Alderson, J.C. and A.H Urquhart (eds.). 1984. *Reading in a foreign language.* Harlow: Longman.

Allen, R.E. (ed.). 1990. *The concise Oxford dictionary of current English.* Oxford: Clarendon Press.

Anderson, I.H. and W.E. Dearborn. 1952. *The psychology of teaching reading.* New York: Ronald Press.

Arnaud, P.J.L. and H. Béjoint (eds.). 1992. *Vocabulary and applied linguistics.* Houndsmills: Macmillan.

Arnold, S., N.Barr, P.J.Donnelly, C.Duffy, P.Gray, and D.Morton, D. 1994. *Constructing and implementing multimedia teaching packages*. Glasgow: University of Glasgow.

Atkinson, R.C. 1974. 'Teaching children to read using a computer'. *American Psychologist.* 29. 169-178.

Bach, M.J. and B.J. Underwood. 1970. 'Developmental changes in memory attributes'. *Journal of Educational Psychology.* 61. 292-296.

Bannon, L.J. 1986. 'Issues in design: some notes'. In D.A. Norman and S.W. Draper (eds). 25-29.

Bates, T. and J. Robinson (eds.). 1977. *Evaluating educational television and radio: proceedings of the international conference on evaluation and research in educational television and radio.* Milton Keynes, England: The Open University.

Bernhardt, E.B. 1991. *Reading development in a second language: theoretical, empirical, and classroom perspectives.* Norwood NJ: Ablex Publishing.

Blair, R.W. (ed.). 1982. *Innovative approaches to language teaching.* Rowley, MA: Newbury House Publishers.

Blair, R.W. 1982. 'An integrated approach'. In R.W. Blair (ed.). 229-243.

Blattner, M.M. 1993. 'Sound in the Multimedia Interface'. In H. Maurer (ed.). 76-82.

Blease, D. 1986. *Evaluating educational software.* London: Croom Helm.

Bonfield, J. 1975. 'Effects of technological change on trade-unions'. *Production Journal.* 66. Jan. 13-14.

Bongaerts, T., P. de Haan, S. Lobbe, and H. Wekker (eds.). 1989. *Computer applications in language learning.* The Netherlands: Foris Publications.

Borg, W.R. and M.D. Gall. 1989. *Educational Research.* New York: Longman.

Borsook, T.K. and N. Higginbotham-Wheat. 1991. 'Interactivity: what is it and what can it do for computer-based instruction'. *Educational Technology.* October 1991. 11-17.

Boyd, G.M. and P.D. Mitchell. 1992. 'How can intelligent CALL better adjust to learners?'. *Computers and Education.* 18/1-3. 23-28.

Brooks, S., Simutis and O'Neil. 1985. 'The role of individual differences in learning strategies research'. In R. Dillon (ed.).

Brown, D.F. 1974. 'Advanced vocabulary teaching: the problem of collocation'. *RELC Journal*. 5/2. 1-11.

Brown, J.D. 1988. *Understanding research in second language learning.* Cambridge: Cambridge University Press.

Business Week (no author named). 1977. 'A technological strike at ABC'. *Business Week*. 2495/ Aug. 8. 67-70.

Business Week (no author named). 1979. 'UAW fears automation again'. *Business Week.* March 26. 94-95.

Business Week (no author named). 1982. 'How to conquer fear of computers'. *Business Week*. 2732/Mar 29. 176-178.

Burt, M.K., H.C. Dulay and M. Finocchiaro (eds.). 1977. *Viewpoints on English as a second language*. New York: Regents Publishing.

Butler, C. (ed.).1992. *Computers and written texts.* Oxford: Blackwell.

Byrum, D.C. 1992. 'Formative evaluation of computer courseware: An experimental comparison of two methods'. *Journal of Educational Computing Research.* 8/1. 69-80.

Carroll, J.M. and R.L. Mack. 1984. 'Learning to use a word processor: by doing, by thinking, and by knowing'. In J.C. Thomas and M.L. Schneider (eds.). 13-52.

Carter, R. 1987. *Vocabulary.* London: Allen and Unwin.

Carter, R. and M. McCarthy (eds.).1988. *Vocabulary and language teaching.* New York: Longman.

Channell, J. 1988. 'Psycholinguistic considerations in the study of L2 vocabulary acquisition'. In R. Carter and M. McCarthy (eds.). 83-96.

Chomsky, C. 1990. 'Books on Videodisc: Computers, video and reading aloud to children'. In D. Nix and R. Spiro (eds.). 31-47.

Clarke, D.F. and I.S.P Nation. 1980. 'Guessing the meaning of words from context: strategy and techniques'. *System.* 8. 211-220.

Coady, J., J. Magoto, P. Hubbard, J. Graney, and K. Mokhtari. 1993. 'High frequency vocabulary and reading proficiency in ESL readers'. In T. Huckin, M. Haynes, and J. Coady (eds.). 217-226.

Cockroft, D. 1980. 'New office technology and employment'. *International Labour Review.* 119/6. 689-704.

Cohen, A.D. 1987. 'Studying learner strategies: how we get the information'. In Wenden and Rubin (eds.). 31-40.

Cohen, A.D. and Hosenfield, C. 1981. 'Some use of mentalistic data in second language research'. *Language Learning.* 31/2. 285-313.

Coltheart, M. 1987. 'Functional architecture of the language-processing system'. In M. Coltheart, G. Sartori, and R. Job (eds.). 1-25.

Coltheart, M., G. Sartori, and R. Job (eds.). 1987. *The Cognitive Neuropsychology of Language.* Hillsdale NJ: Lawrence Erlbaum Associates.

Conrad, R. 1972. 'Speech and reading'. In Kavanagh and Mattingly. (eds.).

Cook, V. (ed.). 1986. *Experimental Approaches to Second Language Learning.* Oxford: Pergamon.

Cook, V. 1986. 'Do second language learners have a cognitive defect?'. In V. Cook (ed.). 73-79.

Corder, S.P. 1967. 'The significance of learners' errors'. *IRAL*. 4. 161-169.

Crable, E.A., J.D. Brodzinski, R.F. Scherer, and P.D. Jones. 1994. 'The impact of cognitive appraisal, locus of control, and level of exposure on the computer anxiety of novice computer users'. *Journal of Educational Computing Research*. 10/4. 329-340.

Craik, F.I.M. and E. Tulving. 1975. 'Depth of processing and the retention of words in episodic memory'. *Journal of Experimental Psychology*. 104/3. 268-294.

Cumming, G., S. Cropp and R. Sussex. 1994. 'On-line lexical resources for language learners: assessment of some approaches to word definition'. *System*. 22/3. 369-377.

Cutler, A. 1983. 'Lexical complexity and sentence processing'. In G.B. Flores d'Arcais and R.J. Jarvella (eds.). 43-79.

Danesi, M. and A. Mollica,. 1988. 'From right to left: a 'bimodal' perspective of language teaching'. *Canadian Modern Language Review*. 45/1. 76-89.

Davidson, J., D.Coles, P.Noyes and C.Terrell. 1991. 'Using computer-delivered natural speech to assist in the teaching of reading'. *British Journal of Educational Technology*. 22/2. 110-118.

Davis, F.B. 1968. 'Research in comprehension and reading'. *Reading Research Quarterly*. 4. 628-678.

Davis, F.B. 1972. 'Psychometric research on comprehension in reading'. *Reading Research Quarterly*. 7. 628-678.

Davies, A. 1992. Personal Communication. University of Edinburgh.

Day, R.R., C. Omura and M. Hiramatsu. 1991. 'Incidental EFL vocabulary learning and reading'. *Reading in a Foreign Language*. 7/2. 541-547.

Deegan, M.. 1992. 'Guide'. *The CTISS File*. No. 13. 27-29.

Deese, J. 1965. *The structure of associations in language and thought*. Baltimore, MD: The Johns Hopkins University Press.

Dhaif, H. 1990. 'Reading aloud for comprehension: a neglected teaching aid'. *Reading in a Foreign Language*. 7/1. 457-46.

Dickinson, L. 1987. *Self instruction in language learning.* Cambridge: Cambridge University Press.

Dillon, R. (ed.) 1985. *Individual differences in cognition: Vol. 2.* New York: Academic Press.

Dulay, H., M. Burt, and S.D. Krashen. 1982. *Language two.* New York: Oxford University Press.

Dwyer, E. 1992. Personal communication. University of Edinburgh.

Edfelt, A.W. 1959. *Silent speech and silent reading.* Stockholm: Almqvist and Wiksell.

Ehrlich, K. and E. Soloway. 1984. 'An empirical investigation of the tacit plan knowledge in programming'. In J.C. Thomas and M.L. Schneider (eds.). 13-134.

Elley, W.B. 1989. 'Vocabulary acquisition from listening to stories'. *Reading Research Quarterly.* 24/2. 174-18.

Ellis, R. 1992. *Second language acquisition and language pedagogy.* Clevedon: Multilingual Matters.

Fay, D. and A. Cutler. 1977. 'Malapropisms and the structure of the mental lexicon'. *Linguistic Inquiry.* 8/3. 505-520.

Flagg, B.N. 1990. *Formative evaluation for educational technologies.* Hillsdale NJ: Lawrence Erlbaum Associates.

Flores d'Arcais, G.B., and R.J. Jarvella (eds.). 1983. *The process of language understanding.* Chichester: John Wiley and Sons.

Forlano, G. and M.N.H. Hoffman. 1937. 'Guessing and telling methods in learning words of a foreign language'. *Journal of Educational Psychology.* 28. 632-636.

Foulke, E. and T.G. Sticht .1969. 'A review of research on the intelligibility and comprehension of accelerated speech'. *Psychology Bulletin.* 72. 50-62.

Fries, C.C. 1945. *Teaching and learning English as a foreign language.* Ann Arbor: University of Michigan Press.

Gairns, R. and S. Redman. 1986. *Working with words: a guide to teaching and learning vocabulary.* Cambridge: Cambridge University Press.

Galletly, J., W. Butcher and S. Daryanani. 1992. 'Hypertext in cognate language learning'. *Journal of Computer Assisted Learning.* 8/1. 25-36.

Goertzen, P.G. 1993. 'Developing hypermedia materials for language learning: vocabulary acquisition through listening to the computer.' Paper given at the EuroCALL '93 conference. Hull, England.

Goertzen, P.G. and R. Howard. 1995. 'Creating multimedia courseware for ESP'. *Edinburgh Working Papers in Applied Linguistics.* 6. 18-26.

Grant, N. 1987. *Making the most of your textbook.* Longman: London.

Green, J.M. 1993. 'Student attitudes toward communicative and non-communicative activities: do enjoyment and effectiveness go together?' *The Modern Language Journal.* 77/1. 1-10.

Green, J.M. and R. Oxford. 1995. 'A closer look at learning strategies, L2 proficiency, and gender'. *TESOL Quarterly.* 29/2. 261-301.

Guide 3.1. 1992. OWL International.

Halliday, M.A.K., A. McIntosh, and P.D. Strevens. 1964. *The linguistic sciences and language teaching.* London: Longmans.

Hammond, N. and L. Allinson. 1989. 'Extending hypertext for learning: An investigation of access and guidance tools'. In A. Sutcliffe and A. Macaulay (eds.). 293-304.

Hardman, L. 1990. 'Introduction to hypertext and hypermedia'. *The CTISS File.* No. 9. 2-6.

Hartley, B. and R. Hasan. 1976. *Streamline: departures.* Oxford: Oxford University Press.

Hatch, E. and A.Lazaraton. 1991. *The research manual: design and statistics for applied linguistics.* New York: Newbury House.

Hatch, E.M. (ed.). 1978. *Second language acquisition: a book of readings.* Rowley, MA: Newbury House.

Hazari, S.I. and R.R. Reaves. 1994. 'Student preferences toward microcomputer user interfaces'. *Computers and Education.* 22/3. 225-229.

Haynes, M. 1993. 'Patterns and perils of guessing in second language reading'. In Huckin *et al.* 46-64.

Henning, G.H. 1974. 'Remembering foreign language vocabulary: acoustic and semantic parameters'. *Language Learning*. 23/2. 185-196.

Higgins, J. and T. Johns. 1984. *Computers in language learning*. London: Collins.

Hilton, J.B. 1974. *Language teaching: a systems approach*. London: Methuen Educational.

Honeyfield, J. 1977. 'Word frequency and the importance of context in vocabulary learning'. *RELC Journal*. 8/2. 35-42.

Hornby, A.S. and A.P. Cowie (eds.).*Oxford advanced learner's dictionary of current English*. 1992 (4th edition). Oxford: Oxford University Press.

Horney, M. 1993. 'Case studies of navigational patterns in constructive hypertext'. *Computers and Education*. 20/3. 257-270.

Howatt, A.P.R. 1984. *A history of English language teaching*. Oxford: Oxford University Press.

Howe, I and W. Howe (eds). 1982. *Short Shorts: An Anthology of the Shortest Stories*. Toronto: Bantam.

Huckin, T., M. Haynes, and J. Coady (eds.). 1993. *Second language reading and vocabulary learning*. Norwood NJ: Ablex.

Huckin, T., and M.Haynes. 1993. 'Summary and future directions'. In T. Huckin, M. Haynes, and J. Coady (eds.). 289-298.

Hulstijn, J.H. 1992. 'Retention in inferred and given word meanings: experiments in incidental vocabulary learning'. In P.J.L. Arnaud and H. Bejoint (eds.). 113-125.

Hulstijn, J.H. 1993. 'When do foreign language readers look up the meaning of unfamiliar words? The influence of task and learner variables'. *The Modern Language Journal*. 77/2. 139-147.

Hult, S., M. Kalaja, O.Lassila and T.Lehtisalo. 1990. 'Hyperreader—an interactive course in reading comprehension'. *System*. 18/2. 189-198.

Ingraham, B. and C. Emery. 1991. ''France Interactive': a hypermedia approach to language training'. *Educational Training and Technology International*. 28/4. 321-33.

Inoue, M. 1994. Personal communication. University of Edinburgh.

Jacobs, G. 1992. 'Hypermedia and discovery-based learning: a historical perspective'. *British Journal of Educational Technology*. 23/2. 113-121.

James, A., and T. Westney (eds.). 1981. *New linguistic impulses in foreign language teaching*. Tubingen: Gunter Narr Verlag.

James, J. 1981. 'Learner variation and learner strategies'. In A. James and P. Westney (eds.). 229-238.

Jenkins, R.J., M.L. Stein and K. Wysocki. 1984. 'Learning vocabulary through reading'. *American Educational Research Journal*. 21/4. 767-787.

Jensen, A.R., and W.D. Rohwer. 1965. 'Syntactical mediation of serial and paired-associate learning as a function of age'. *Child Development*. 36. 601-608.

Johns, T. 1986. 'Micro-Concord: a language learner's research tool'. *System*. 14/2. 151-162.

Johns. T. 1989. 'Whence and wither classroom concordancing?'. In Bongaerts (ed.).

Jonassen, D.H. 1990. 'Semantic network elicitation: tools for structuring hypertext'. In R. McAleese and C. Green (eds.). 142-152.

Kang, S. 1995. 'The effects of a context-embedded approach to second-language vocabulary learning'. *System*. 23/1. 43-55.

Kavanagh, J.F. and I.G. Mattingly. (eds.). 1972. *Language by eye and ear*. Cambridge, Mass. M.I.T. Press.

Kinsella, V. (ed.). 1982. *Cambridge language teaching surveys 1*. Cambridge: Cambridge University Press.

Krashen, S.D. 1977. *The monitor model for adult second language performance*. In M.K. Burt, H.C. Dulay and M. Finocchiaro (eds.). 152-161.

Krashen, S.D. 1989. 'We acquire vocabulary and spelling by reading: additional evidence for the input hypothesis'. *Modern Language Journal*. 73/4. 441-46.

Krendl, K.A. and M. Broihier. 1992. 'Student responses to computers: a longitudinal study'. *Journal of Educational Computing Research*. 8/2. 215-227.

Lado, R. 1961. *Language testing: the construction and use of foreign language tests*. London: Longman.

Laufer, B. 1986a. 'Possible changes in attitude towards vocabulary acquisition research'. *IRAL*. 24/1. 69-75.

Laufer, B. 1986b. *Vocabulary acquisition in a second language: the hypothesis of 'synforms' (similar lexical forms).*Unpublished Ph.D. Thesis. University of Edinburgh.

Laufer, B. 1990. 'How much lexis is necessary for reading comprehension?'. In P.J.L Arnaud and H. Béjoint (eds.). 126-132.

Laufer, B. 1991. 'The development of L2 lexis in the expression of the advanced learner'. *Modern Language Journal.* 75/4. 440-448.

Laurel, B.K. 1986. 'Interface as memesis'. In D.A. Norman and S.W. Draper (eds.). 67-85.

Leffa, V.J. 1993. 'Making foreign language texts comprehensible for beginners.' *System.* 20/1. 63-73.

Legenhausen, L. and D. Wolff. 1990. 'CALL in use--use of CALL: evaluating CALL software. *System.* 18/1. 1-13.

Lightbown, P.M. and Pienemann, M. 1993. 'Formal grammar instruction: *TESOL Quarterly.* 27/4. 717-722.

Malone, T.W. 1984. 'Heuristics for designing enjoyable user interfaces: lessons from computer games'. In J.C. Thomas and M.L. Schneider (eds.). 1-12.

Mark, W. 1986. 'Knowledge-based interface design.' In D.A. Norman and S.W. Draper (eds.). 219-242.

May, C. 1986. *Effects of reading aloud in English on the reading ability and attitudes of Spanish-speaking children.* Unpublished Ph.D. Thesis. University of Massachusetts.

Maurer, H. (ed.). 1993. *International multimedia and hypermedia annual, 1993: proceedings of Ed-Media 93—World conference on educational multimedia and hypermedia.* Orlando FL: Association for the Advancement of Computing in Education.

McAleese, R. and C. Green (eds.). 1990. *Hypertext: state of the art.* Oxford: Intellect.

McCarthy, M.J. 1984. 'A new look at vocabulary in EFL'. *Applied Linguistics.* 5/1. 12-22.

McInerney, V., D.M. McInerney and K.E. Sinclair. 1994. 'Student teachers, computer anxiety and computer experience'. *Journal of Educational Computing Research.* 11/1. 27-50.

McKnight, C., A. Dillon and J. Richardson. 1991. *Hypertext in context.* Cambridge: Cambridge University Press.

Meara, P. 1982. 'Vocabulary acquisition: a neglected aspect of language learning'. In V. Kinsella (ed.). 100-126.

Meara, Paul. 1989. 'Matrix Models of Vocabulary Acquisition'. *AILA Review*. Vol.6. 66-74.

Meara, P. 1986. 'Psycholinguistic relativity—is Spanish different?'. In V. Cook (ed.). 151-156.

Meara, P. 1990. 'A note on passive vocabulary'. *Second Language Research*. 6/2. 150-154.

Meara, P. 1992. 'Network structures and vocabulary acquisition in a foreign language'. In P.J.L Arnaud and H. Bejoint (eds). 62-70.

Meara, P. and G. Jones. 1989. *Eurocentres vocabulary test 10 KA*. Zurich: Eurocentres.

Mondria, J.A. and M. Wit-De Boer. 1991. 'The effects of contextual richness on the guessability and the retention of words in a foreign language'. *Applied Linguistics*. 12, 3. 249-67.

Mouly, G.J. 1978. *Educational research: the art and science of investigation.* Boston: Allyn and Bacon.

Nagy, W.E., R.C. Anderson and P.A. Herman. 'Learning word meanings from context during normal reading'. *American Educational Research Journal.* 24/2. 237-270.

Nation, I.S.P and J. Coady. 1988. 'Vocabulary and reading'. In R. Carter and M. McCarthy (eds.). 83-96.

Nation, I.S.P. 1982. 'Beginning to learn foreign vocabulary: a review of the research'. *RELC Journal*. 13/1. 15-36.

Nation, I.S.P. 1983. *Learning and teaching vocabulary.* Wellington New Zealand: Victoria University.

Nesi, H. and P. Meara. 1994. 'Patterns of misinterpretation in the productive use of EFL dictionary definitions'. *System*. 22/1. 1-15.

Neville, M.H. 1968. 'Effects of oral and echoic responses in beginning reading'. *Journal of Educational Psychology*. 59. 362-369.

Neville, M.H. and A.K. Pugh. 1974. 'An empirical study of the reading while listening method'. Paper presented at *The Eleventh Annual Conference of the United Kingdom Reading Association.*

Nickerson, R.S. 1986. *Using computers: human factors in information systems*. Cambridge MA: MIT Press.

Nix, D. and R. Spiro (eds.). 1990. *Cognition, education and multimedia: exploring ideas in high technology*. Hillsdale NJ: Lawrence Erlbaum.

Norman, D.A., and S.W. Draper (eds.). 1986. *User centered system design: new perspectives on human-computer interaction*. Hillsdale NJ: Lawrence Erlbaum.

Norman, D.A., and D.E. Rumelhart (eds.). 1975. *Explorations in cognition*. San Francisco: W.H. Freeman and Co.

O'Malley, J.M., A.U. Chamot, G. Stewner-Manzaneres, L. Kupper, and R.P. Russo. 1985. 'Learning strategies used by beginning and intermediate ESL students'. *Language Learning*. 35/1. 21-46.

O'Malley, J.M. and A.U. Chamot. 1990. *Learning Strategies in Second Language Acquisition*. Cambridge: Cambridge University Press.

O'Neill, R. 1978. *Kernel one*. London: Longman.

Oettinger, A.G. 1969. *Run, computer run: the mythology of educational innovation*. Harvard MA: Harvard University Press.

Ott, C.E., S.Blake, and J.P.Ball. 1974. 'The effect of interactive-image elaboration on the acquisition of foreign language vocabulary'. *Language Learning*. 23/2. 197-206.

Oxford, R.L. and R.C. Scarcella. 1994. 'Second language vocabulary learning among adults: state of the art in vocabulary instruction'. *System*. 22/2. 231-243.

Paz, O. 1982. 'The blue bouquet'. In Howe, I and I.W. Howe (eds). 163-165.

Pennington, M.C. and V. Stevens (eds). 1992. *Computers in Applied linguistics*. Clevedon: Multilingual Matters.

Pitts, M., H. White, and S.D. Krashen. 1989. 'Acquiring second language vocabulary through reading: a replication of the clockwork orange study using second language acquirers'. *Reading in a Foreign Language*. 5/2. 271-275.

Reber, A.S., F.F. Walkenfield, and R. Hernstadt. 1991. 'Implicit and explicit learning: individual differences and IQ'. *Journal of Educational Psychology: Learning, Memory and Cognition*. 17/5. 888-896.

Reid, S.L. 1971. 'The effect on reading achievement of reading paced by compressed speech. Bloominton: Indiana University.

Reilard, M.K. 1970. 'The effect of compressed speech on reading and listening abilities of sixth-grade children'. New Brunswick: Rutgers University.

Quentin-Baxter, M. and D. Dewhurst. 1992. 'A method for evaluating the efficiency of presenting information in a hypermedia environment'. *Computers and Education*. 18/1-3. 179-182.

Rasinski, T.V. 1990. 'Effects of repeated reading and listening-while-reading on reading fluency'. *Journal of Educational Research*. 83/3. 147-150.

Reitsma, P. 1988. 'Reading practice for beginners: effects of guided reading, reading-while-listening, and independent reading'. *Reading Research Quarterly*. 23/2. 219-235.

Richards, J.C. 1974. 'Word lists: problems and prospects'. *RELC Journal*. 5/2. 69-84.

Richards, J.C. 1976. 'The role of vocabulary teaching'. *TESOL Quarterly*. 10/1. 77-89.

Richmond. W.K. 1965. *Teachers and machines: an introduction to the theory and practice of programmed learning*. London: Collins.

Rivers, W.M. 1968. *Teaching foreign language skills*. Chicago: University of Chicago Press.

Rowntree, D. 1981. *Statistics without tears*. London: Penguin.

Rumelhart, D.E. and D.A.Norman. 1975. *Explorations in Cognition*. San Fransisco: Freeman.

Santos, J. 1987. *The effects of the teacher's reading aloud on the second language learner's reading comprehension, listening comprehension, and attitude towards reading*. Unpublished Ph.D. Thesis. Boston University.

Saragi, T., I.S.P. Nation and G.F. Meister. 1978. 'Vocabulary learning and reading'. *System*. 6/72. 72-78.

Schouten-Van Parreren, M.C. 1985. *Woorden leren in het vreemde-talenonderwijs.* Apeldoorn: Van Walraven.

Schramm, W. 1977. Foreword. In T.Bates and J.Robinson (eds.). vii-viii.

Sciarone, A.G. and P.J. Meijer. 1993. 'How free should students be? A case from CALL: computer-assisted language learning'. *Computers and Education*. 21/1,2. 95-101.

Selinker, L. 1972. 'Interlanguage'. *IRAL*. 10. 209-231.

Shneiderman, B. 1987. *Designing the user interface*. Reading MA: Addison Wesley.

Shu, H., R.C. Anderson, and H. Zhang. 1995. 'Incidental learning of word meanings while reading: A Chinese and American cross-cultural study'. *Reading Research Quarterly*. 30/1. 76-95.

Shwalb, B. 1994. 'A review of research comparing computer-assisted instructions and traditional instructions'. *Journal of Communication and International Studies*. 1/1. 37-42.

Skinner, B.F. 1954. 'The science of learning and the art of teaching'. *Harvard Educational Review*. 24/2. 86-97.

Sokolov, A.N. 1960. 'Silent speech in the study of foreign languages'. *Voprosy Psichologii*. 5. 57-64.

Sonaiya, R. 1991. 'Vocabulary acquisition as a process of continuous lexical disambiguation'. *IRAL*.29/4. 273-284.

SoundBlaster 16. 1992. Creative Research Inc.

Spearitt. 1972. 'Identification of subskills of reading: comprehension by maximum likelihood factor analysis. *Reading Research Quarterly*. 8. 92-111.

Spiro, R., and J. Jehng. 1990. 'Cognitive flexibility and hypertext: theory and technology for the nonlinear and multidemensional traversal of complex subject matter'. In D. Nix and R. Spiro (eds.). 163-204.

Stanton, N. and R.B. Stammers. 1990. 'Learning styles in a non-linear training environment'. In R. McAleese and C. Green (eds.).114-120.

Stevens, V. 1991. 'A study of student attitudes toward CALL in a self-access student resource centre'. *System*. 19/3. 289-299.

Stevens, V. 1991b. 'Classroom concordancing: vocabulary materials derived from authentic, relevant text'. *English for Specific Purposes*. 10. 35-46.

Stevens, V. 1992. 'Humanism and CALL: a coming of age'. In Pennington and Stevens (eds.).11-38.

Sudman, S. and N.M. Bradburn. 1982. *Asking questions: a practical guide to questionnaire design*. San Fransisco: Jossey-Bass Publishers.

Sutcliffe, A. and A. Macaulay (eds.). 1989. *People and computers V.* Oxford: Oxford University Press.

Tarone, E. 1974. Speech perception in second language acquisition: a suggested model. *Language Learning*. 24: 223-233.

Thomas, J.C. and M.L. Schneider (eds.). 1984. *Human factors in computer systems*. Norwood NJ: Ablex.

Tessler, F.N. and H. Singh. 1995. 'SuperCard 1.7.1'. *Macworld*. February. 68.

Twaddell, F. 1973. 'Vocabulary expansion in the TESOL classroom'. *TESOL Quarterly*. 7. 61-78.

Underwood, J.D.M. and G. Underwood. 1990. *Computers and learning: helping children acquire thinking skills*. Oxford: Basil Blackwell.

Unwin, D. and R. McAleese (eds.). 1978. *Encyclopaedia of educational media communications and technology*. Westport, CT: Greenwood Press.

van der Linden, E. 1993. 'Does feedback enhance computer-assisted language learning?'. *Computers and Education*. 21/1,2. 61-65.

Wendin, A. and J. Rubin. (eds.).1987. *Learner Strategies in Language Learning*. Englewood Cliffs, NJ: Prentice-Hall.

West, M. 1953. *A general service list of English words*. London: Longman.

Whistlecroft, L. 1990. 'Hypercard'. *CTISS File*. 9. 20-23.

Widdowson, H.G. and C.J. Brumfit. 1981. 'Issues in second language syllabus design'. In J.E. Alatis, H.B. Altman and P.M. Alatis (eds.). 199-210.

Williams, N. 1992. 'Computers and writing'. In C. Butler (ed.). 247-265.

Wittrock, M.C., C.B. Marks, and M.J. Doctorow. 1975. 'Reading as a generative process'. *Journal of Educational Psychology*. 67. 484-489.

Wise, B, R. Olson, M. Ansett, L. Andrews, M.Terjak, V. Schneider, J. Kostuch, and L. Kriho. 1989. 'Implementing a long-term computerised remedial reading program with synthetic speech feedback: hardware, software and real-world issues'. *Behaviour Research Methods, Instruments and Computers.* 21. 173-171.

Wode, H. 1981. *Learning a second language: an integrated view of language acquisition.* Tubingen Germany: Gunter Narr Verlag.

Woodhead, N. 1990. *Hypertext and hypermedia: theory and applications.* Wilmslow England: Sigma Press.

Yap, K.O. 1979. 'Vocabulary—building blocks of comprehension?'. *Journal of Reading Behavior.* 1. 49-59.

Appendices

Table of Contents: Appendix

Appendix 1

Appendix 1 .1 : Overview of Hypermedia Software

HyperCard

HyperCard is one of the original hypermedia programs. It was created to fill a perceived niche in the industry, i.e., a programming environment that almost anybody could use. In 1987, when HyperCard became commercially available (it was 'bundled' with every Mac sold in North America), existing programming languages were powerful but prohibitively difficult for the majority of computer owners. The intuitively simple interface of the Macintosh was difficult for even advanced programmers to create software for. In the IBM world, simple programming languages like BASIC were easy to learn but not powerful or flexible enough. The creator of HyperCard (Bill Atkinson) wanted something that would provide "programming for the masses" or a "software erector set". (Hult *et al.* 1990). Continuing this metaphor, one can imagine a program consisting of standardised elements (or building blocks) which an author can manipulate to suit his or her purposes. The existence of flexible, pre-fabricated components was a precursor for all hypermedia programs and remains a prerequisite for any viable hypermedia application.

HyperCard treats the computer screen like a stack of index cards. These cards can be sorted, shuffled, linked to other cards or even animated. Hypercard is equipped with a simple drawing program and can also use 'clip-art' images available from many public domain archives. The index card metaphor is particularly suited to displaying screens of information accompanied by a picture or explanation. The author can add buttons or hotspots to the cards.

The strength of HyperCard lies in its built-in programming language called Hypertalk. Hypertalk is simple. Commands are very similar to everyday English and the program is amazingly tolerant of mistakes in syntax and spelling (Whistlecroft 1990). Most interestingly for the language teacher, text can be accompanied by pre-recorded, digitised sound. To record sounds, most Macs are equipped with a small microphone and a software application to play them back. Sounds can be linked to words or pictures creating a variety of possibilities for language instruction.

Toolbook

Toolbook could be called the PC version of Apple's Hypercard; it includes its own programming language and also offers simple routines that can be organised to make elaborate textual and graphical documents. It provides a good shell for classroom use (using a large screen computer or an overhead projector linked to a PC) because pages on the screen can be made automatically to appear consecutively. Toolbook would be an ideal program for 'mixed shops'—schools or offices that have both Apple and IBM machines, because it conforms to an industry standard called HIFF (Hypertext

Interchange File Format) which allows Toolbook documents to be converted to and from Hypercard.

Guide

Guide, being one of the first hypermedia programs for DOS computers, enjoys a fairly wide user base—though not as wide as HyperCard. Guide shares many of the elements of HyperCard including easy creation of buttons, and the ability to efficiently include graphics with text. Following the HyperCard model, Guide has a built in programming language called Logiix. Logiix is based on Pascal, an extremely powerful tool for expanding the capability of a hypermedia document. Unfortunately, Logiix is not as easy to use as Hypertalk and the documentation, though thoroughly describing the commands of the programming language, fails to provide enough practical examples for things such as creating dialogue boxes or making links to sound elements.

Guide lends itself well to text-based documents (Deegan 1992). There are several ways in which an author can give explanations and expansions on text or graphics. All the fonts possible in the Windows environment (including foreign alphabets— even biblical Hebrew) are available for the document designer. Given its strength in this area, Guide would be an ideal platform for lessons designed to practice and teach reading or writing. It is possible to create links to digitised sound and also to run other programs and peripheral devices like CD ROM players from within Guide. Guide includes an authoring program and a reader program; this is beneficial in two ways: the user is unable to destroy the document and the reader program may be freely distributed without additional charge.

There are several separate applications available for Guide that might be of use to a language teacher. The *Guide Video Toolkit* allows video to be cut and pasted into a Guide document using a video sequence editor coupled with a playback interpreter (Deegan). This would make possible motion video sequences that might be used to present English in situations, with the usual hypermedia possibilities of feedback, hints, questions, and summarisation of video events. Another utility for Guide, called *Guidebook,* allows Guide documents to be run in Asymetrix Toolbook. Any program that makes provision for other programs and standards (especially with competitor's products) must be commended. Open systems allow for greater portability and less frustration for designers.

Visual Basic

BASIC was a programming language initially supplied free with every IBM compatible. The development of Windows has lead to a concurrent change in the interface of BASIC. It has now become a graphically oriented program allowing the programmer to create on-screen buttons, text boxes and other system controls simply by drawing them on the screen and attaching them with code to various tasks. The obstacle, of course, is the code. It is impossible to do anything with Visual Basic without writing computer code. The code is very similar to traditional BASIC but nevertheless requires a great deal of time to learn. Visual Basic, in the professional version, is capable of hypermedia features like playing digitised sound, MIDI files, Microsoft Video, and animation. The advantage of Visual Basic (and any other programming language), is that it can produce free-standing programs that only require a small 'run-time' library (containing various pre-programmed routines). Perhaps the greatest advantage of free-standing programs is that they are owned by the programmer and require no licence fees to other parties. It would thus be an appropriate tool for developers wishing to market their creations or schools that cannot afford multiple copies of programs such as Toolbook. No doubt, the growing support in industry for Visual Basic and the fact that most Microsoft programs make use of it in some form will lead to a greater use of Visual Basic among language teachers in the future. This researcher has found it to be an extremely flexible and useful tool. It is his authoring tool of choice.

Icon Author

Icon author is reported to be a powerful, functional, and extremely flexible hypertext/hypermedia authoring tool. It comes with example applications and is well-supported by on-line help, readable user manuals and a 'getting started' book. There is one essential ingredient missing in this program, however—an affordable price. The suggested retail price for Icon Author is £2595.00 plus VAT and the educational price is more than £800.00. The company clearly is aiming its software at large companies or software developers as it is well beyond the reach of most language schools or universities.

Microsoft Access

To anyone familiar with the Microsoft Office range of products, the notion of using Access for language learning should come as a surprise — it is primarily considered a business database. However, we have used Access with a large degree of success, creating case problems for Medical English (Goertzen and Howard 1995) and developing 'information browsers' for students at the Institute for Applied Language Studies at the University of Edinburgh. Access works on most Windows computers. However, a fast 486 computer with 8 megabytes of RAM and a large hard disk is required to avoid frustration in the design phase. The advantage of Access is that very little programming is required to manipulate screens (in Access they are called 'forms'). Keeping track of users' names, actions, navigational patterns, etc., is fairly easy as the database table can be used without any programming whatsoever. Access works flawlessly with programs developed in Visual Basic. Indeed, most of the Visual Basic programming language is built in to Access. Learning one is learning the other.

Summary

The prospective user of hypermedia software is thus faced with a fairly broad choice. Not listed here are a number of programs dedicated to single tasks such as reading and writing. Programs such as this are available commercially, or it is possible to modify existing programs like HyperCard to act as templates for courseware development. Hypermedia development software must have a minimum capability for practical use in language teaching. The following table summarises the requirements of such software:

- can link to digitised sound files
- no inordinate sense of waiting
- possibility of producing templates
- simple graphic insertion and utilisation
- extendible to hypermedia peripherals
- separate programming capability
- simplicity of programming language
- intuitive interface

Appendix 2

Appendix 2 .1 : Transcript of The Blue Bouquet

The Blue Bouquet
by Octavio Paz

I woke covered with sweat. Hot steam rose from the newly sprayed, red-brick pavement. A grey-winged butterfly, dazzled, circled the yellow light. I jumped from my hammock and crossed the room barefoot, careful not to step on some scorpion leaving his hideout for a bit of fresh air. I went to the little window and inhaled the country air. One could hear the breathing of the night, feminine, enormous. I returned to the center of the room, emptied water from a jar into a pewter basin, and wet my towel. I rubbed my chest and legs with the soaked cloth, dried myself a little, and, making sure that no bugs were hidden in the fold of my clothes, got dressed. I ran down the green stairway. At the door of the boardinghouse I bumped into the owner, a one-eyed taciturn fellow. Sitting on a wicker stool, he smoked, his eye half closed. In a hoarse voice, he asked:

"Where are you going?"

"To take a walk. It's too hot."

"Hmmm--everything's closed. And no streetlights around here. You'd better stay put." I shrugged my shoulders, muttered "back soon," and plunged into the darkness. At first I couldn't see anything. I fumbled along the cobblestone street. I lit a cigarette. Suddenly the moon appeared from behind a black cloud, lighting a white wall that was crumbled in places. I stopped, blinded by such whiteness. Wind whistled slightly. I breathed the air of the tamarinds. The night hummed, full of leaves and insects. Crickets bivouacked in the tall grass. I raised my head: up there the stars too had set up camp. I thought that the universe was a vast system of signs, a conversation between giant beings. My actions, the cricket's saw, the star's blink, were nothing but pauses and syllables, scattered phrases from that dialogue. What word could it be, of which I was only a syllable? Who speaks the word? To whom is it spoken? I threw my cigarette down on the sidewalk. Falling, it drew a shining curve, shooting out brief sparks like a tiny comet.

I walked a long time, slowly. I felt free, secure between the lips that were at that moment speaking me with such happiness. The night was a garden of eyes. As I crossed the street, I heard someone come out of a doorway. I turned around, but could not distinguish anything. I hurried on. A few moments later I heard the dull shuffle of sandals on the hot stone. I didn't want to turn around, although I felt the shadow getting closer with every step. I tried to run. I couldn't. Suddenly I stopped short. Before I could defend myself, I felt the point of a knife in my back and a sweet voice;

"Don't move, mister, or I'll stick it in."

Without turning, I asked;

"What do you want?"

"Your eyes, mister," answered the soft, almost painful voice.

"My eyes? What do you want with my eyes? Look, I've got some money. Not much, but it's something. I'll give you everything I have if you let me go. Don't kill me."

"Don't be afraid, mister, I won't kill you. I'm only going to take your eyes."

"But why do you want my eyes?" I asked again.

"My girlfriend has this whim. She wants a bouquet of blue eyes. And around here they're hard to find."

"My eyes won't help you. They're brown, not blue."

"Don't try to fool me, mister. I know very well that yours are blue."

"Don't take the eyes of a fellow man. I'll give you something else."

"Don't play saint with me," he said harshly. "Turn around."

I turned. He was small and fragile. His palm sombrero covered half his face. In his right hand he held a country machete that shone in the moonlight.

"Let me see your face."

He struck another match, and put it near my eyes. Grabbing my sleeve, he ordered:

"Kneel down."

I knelt. With one hand he grabbed me by the hair, pulling my head back. He bent over me, curious and tense, while his machete slowly dropped until it grazed my eyelids. I closed my eyes.

"Keep them open," he ordered.

I opened my eyes. The flame burned my lashes. All of a sudden he let me go.

"All right, they're not blue. Beat it." He vanished. I leaned against the wall, my head in my hands. I pulled myself together. Stumbling, falling, trying to get up again. I ran for an hour through the deserted town. When I got to the plaza, I saw the owner of the boardinghouse, still sitting in the front of the door. I went in without saying a word. The next day I left town.

Appendix 2 .2 : Sample Log File from the Experimental Group

Note: The text in the file indicates the coded hotspot that the student pressed. The coding is simply the first two words of the sentence combined or the citation form of the word, if it is a word. For analysis, the file would be subsequently converted to a list by using various search and replace methods.

(15:49:48)Iwoke (15:49:53)Hotsteam (15:49:58)AGrey (15:50:12)IJumped (15:50:18)IWent (15:50:24)OneCould (15:50:32)IReturned (15:50:43)IRubbed went ahead one page at: 15:50:43 (15:50:52)IRan (15:50:58)AtTheDoor (15:51:01)taciturn (15:51:06)SittingOn (15:51:13)InAHoarse (15:51:16)WhereAre (15:51:22)ToTake (15:51:27)AndNot (15:51:32)You'dBetter (15:51:38)IShrugged (15:51:42)AtFirst (15:51:50)IFumbled (15:51:55)IFumbled (15:51:59)ILit (15:52:07)SuddenlyThe (15:52:12)IStopped went ahead one page at: 15:52:13 (15:52:18)WindWhistled (15:52:22)IBreathed (15:52:27)TheNight (15:52:31)CricketsBivouacked (15:52:37)IRaised (15:52:43)Upthere (15:52:52)IThought (15:53:02)MyAction (15:53:11)MyAction (15:53:19)WhatWord (15:53:24)WhoSpeaks (15:53:26)ToWhom

(15:53:31)IThrew (15:53:36)FallingIt (15:53:43)IWalked (15:53:51)IFelt (15:53:54)TheNight (15:53:59)AsICrossed went ahead one page at: 15:53:59 (15:54:06)ITurned (15:54:08)IHurried (15:54:14)AFew (15:54:19)IDidn't (15:54:23)ITried (15:54:25)ICouldn't (15:54:28)SuddenlyIStopped (15:54:34)BeforeI (TueNov 23 15:54:38)Don'tMove (15:54:43)WithoutTurning (15:54:48)YourEyes (15:54:52)MyEyes (15:54:56)WhatDo(15:54:59)LookI've went ahead one page at: 15:55:00 (15:55:05)NotMuch (15:55:09)I'llGive (15:55:11)Don'tKill (15:55:15)Don'tBe (15:55:19)I'mOnly (15:55:23)ButWhy (15:55:28)MyGirlfriend (15:55:32)SheWants (15:55:36)AndAround (15:55:38)MyEyes (15:55:42)They'reBrown (15:55:46)Don'tTry (15:55:51)IKnow (15:55:55)Don'tTake (15:55:58)I'llGive (15:56:02)Don'tPlay (15:56:07)TurnAround (15:56:09)ITurned (15:56:13)HeWasSmall went aheadone page at: 15:56:15 (15:56:23)InHisRight (15:56:26)LetMe (15:56:31)HeStruck (15:56:34)GrabbingMy (15:56:40)IKnelt (15:56:44)WithOne (15:56:55)HeBent (15:56:58)IClosed (15:57:01)KeepThem (15:57:13)IOpened (15:57:16)TheFlame (15:57:20)AllOf (15:57:23)AllRight (15:57:31)BeatIt went ahead one page at: 15:57:32 (15:57:36)HeVanished (15:57:41)ILeaned (15:57:44)IPulled (15:57:48)StumblingFalling (15:57:52)IRan (15:57:58)WhenI (15:58:01)IWentIn went back one page at: 15:58:04 went back one page at: 15:58:05 went back one page at: 15:58:05 went back one page at: 15:58:06 went back one page at: 15:58:06 went back one page at: 15:58:08 (15:58:14)Iwoke (15:58:18)Hotsteam (15:58:24)AGrey (15:58:32)IJumped (15:58:36)IWent (15:58:42)OneCould (15:58:49)IReturned (15:59:01)IRubbed (15:59:09)IRubbed went ahead one page at: 15:59:12 (15:59:19)IRan (15:59:25)AtTheDoor (15:59:30)taciturn (15:59:35)SittingOn (15:59:42)SittingOn (15:59:47)InAHoarse (15:59:50)WhereAre (15:59:55)ToTake (16:00:00)AndNot (16:00:07)You'dBetter (16:00:14)IShrugged (16:00:18)AtFirst (16:00:22)IFumbled (16:00:33)SuddenlyThe (Tue Nov23 16:00:37)IStopped went ahead one page at: 16:00:39 (16:00:44)WindWhistled (16:00:48)IBreathed (16:00:52)TheNight (16:00:57)CricketsBivouacked (16:00:59)bivouacked (16:01:06)IRaised (16:01:10)Upthere (16:01:21)IThought (16:01:29)MyAction (16:01:38)MyAction (16:01:42)WhatWord (16:01:46)WhoSpeaks (16:01:50)ToWhom (16:01:54)IThrew (16:02:00)FallingIt (16:02:04)IWalked (16:02:12)IFelt (16:02:19)IFelt (Tue Nov23 16:02:25)TheNight (16:02:34)AsICrossed went ahead one page at: 16:02:34 (16:02:39)ITurned (16:02:42)IHurried (16:02:48)AFew (16:02:54)IDidn't (16:02:57)ITried (16:02:58)ICouldn't (16:03:01)SuddenlyIStopped (16:03:07)BeforeI (TueNov 23 16:03:14)BeforeI (16:03:20)Don'tMove (16:03:26)WithoutTurning (TueNov 23 16:03:31)YourEyes (16:03:33)MyEyes (16:03:36)WhatDo (16:03:42)LookI've went ahead one page at: 16:03:42 (16:03:46)NotMuch (16:03:52)I'llGive (16:03:56)Don'tKill (16:04:00)Don'tBe (16:04:03)I'mOnly (16:04:07)ButWhy (16:04:10)MyGirlfriend (16:04:13)whim (16:04:16)SheWants (16:04:20)AndAround (16:04:23)MyEyes (16:04:28)They'reBrown (16:04:31)Don'tTry (16:04:35)IKnow (16:04:38)Don'tTake (16:04:46)I'llGive (16:04:50)Don'tPlay (16:04:55)TurnAround (16:05:03)HeWasSmall (16:05:05)HisPalm (16:05:09)HisPalm (16:05:11)HisPalm (16:05:13)HisPalm (16:05:15)HisPalm (TueNov 23 16:05:16)HisPalm (16:05:18)HisPalm went ahead one page at: 16:05:23 (16:05:31)InHisRight (16:05:37)InHisRight (16:05:40)LetMe (16:05:44)HeStruck (16:05:50)GrabbingMy (16:05:56)WithOne (16:06:04)HeBent (16:06:06)IClosed (16:06:10)KeepThem (16:06:14)IOpened (16:06:17)TheFlame (16:06:25)AllOf (16:06:28)AllRight (16:06:34)BeatIt went ahead one page at: 16:06:35 (16:06:39)HeVanished (16:06:43)ILeaned (16:06:49)ILeaned (16:06:56)IPulled (16:07:00)StumblingFalling (16:07:05)IRan (16:07:12)WhenI (16:07:21)IWentIn(16:07:24)TheNext went back one page at: 23 16:07:26 went back one page at: 16:07:27 went back one page at: 16:07:27 went back one page at: 16:07:28 went back one page at: 16:07:30 went back one page at: 16:07:31 (16:07:49)AGrey (16:08:25)IReturned went ahead one page at: 16:08:39 (16:09:05)shrugged (16:09:10)IShrugged (16:09:16)IFumbled (16:09:23)IFumbled (16:09:37)IStoppedwent ahead one page at: 16:09:37 (16:09:41)WindWhistled (16:09:45)IBreathed (16:09:49)TheNight (16:09:55)CricketsBivouacked (16:09:58)IRaised (16:10:03)Upthere (16:10:13)IThought (16:10:22)MyAction (16:10:31)MyAction (16:10:35)WhatWord (16:10:47)FallingIt (16:10:59)IFelt (16:11:08)AsICrossed went ahead one page at: 16:11:09 (16:11:24)AFew (16:11:32)IDidn't (16:11:43)BeforeI went ahead one page at: 16:11:59 (16:12:14)Don'tBe (16:12:24)whim (16:12:31)bouquet2 (16:12:34)bouquet2 (16:13:04)HeWasSmall went ahead one page at: 16:13:08 (16:13:21)InHisRight (16:13:30)HeStruck THIS CONCLUDES LOG OF: kazuhito.txt

Appendix 2 .3 : Multiple Choice Post-Test

Instructions: Circle the answer that is most like the meaning of the word in the story.
Example:
woke
- a. tent used by wandering desert dwellers
- b. funeral where the relatives of a dead person stay up all night with the body
- c. walk (past tense)
- d. stopped sleeping

1. barefoot
- a. not wearing any shoes
- b. foot of a big furry animal
- c. walk carefully
- d. have very big feet

2. bivouac
- a. decoration worn on the shoulder of certain uniforms
- b. soldiers' camp that doesn't have any tents
- c. weapon used to shoot down airplanes
- d. area not highly populated

3. blink
- a. shine a light unsteadily
- b. look at from above
- c. move very slowly
- d. hit or give a blow to

4. boardinghouse
- a. unused house with wood over the windows
- b. house made of wood
- c. home of a small furry animal
- d. private house where people pay to stay and eat

5. bouquet
- a. food cooked in the French manner
- b. bunch of flowers
- c. tool used for building
- d. small area of water

6. cobblestone
- a. large round stone used for making roads
- b. sound made by geese
- c. stone used as a weapon
- d. stone which sticks up and trips people

7. comet
- a. someone who likes practical jokes
- b. religious ceremony that occurs in spring
- c. star-like object in the sky with a long tail
- d. game that uses racquets and is played on a court

8. crumble
- a. fall or break into small pieces
- b. fold paper many times
- c. fail or do something unsatisfactorily
- d. colour darkly

9. dazzle
- a. colour of leaves in the autumn
- b. dance step
- c. make unable to see clearly
- d. speak harshly

10. deserted
- a. hot sandy place with no trees
- b. food you eat after supper
- c. covered with sand
- d. all the people have left

11. distinguish
  a. see or hear clearly
  b. succeed in being able to do something
  c. delay or stop for a certain length of time
  d. lift up or raise to a higher position

12. enormous
  a. done without planning in advance
  b. very big
  c. thinking too much of yourself or proud
  d. using or needing much effort or strength

13. fold
  a. bend in a piece of clothing
  b. place to hide
  c. female whale
  d. jump up and run away

14. fumble
  a. go quickly, hurriedly
  b. speak in a quiet, unclear voice
  c. mix up
  d. use the hands without skill

15. grab
  a. knock something over clumsily
  b. take something quickly
  c. tie a knot in
  d. stupid or ridiculous

16. graze
  a. wire fence used on a farm
  b. touch lightly in passing
  c. take more than your share
  d. flower arrangement that hangs on the wall

17. grey-winged
  a. cowardly
  b. with grey coloured wings
  c. not human, animal like
  d. unclear

18. hammock
  a. muscle in the back of your leg
  b. kind of fish in tropical oceans
  c. bed made of cloth and ropes hung between two posts
  d. ball used in water polo

19. hideout
  a. place to hide or escape to
  b. hide outside
  c. go outside and stay outside
  d. hide with other people

20. Hmmm
  a. bright light that always stays on
  b. kind of bamboo window
  c. sound you make when you're thinking
  d. tropical insect that bites, like a mosquito

21. **hoarse**
   a. yellow flying insect
   b. four legged animal you can ride
   c. play roughly with another person
   d. rough and harsh sounding

22. **hum**
   a. just miss the target
   b. make a sound like the letter *m* for a long time
   c. go or move slowly like "slow motion"
   d. fail to do something correctly

23. **inhale**
   a. slowly destroy
   b. snow-like ice that falls like rain
   c. breath in
   d. greet someone kindly

24. **lashes**
   a. race for small sailing boats
   b. small plant-eating animals who live in warm climates
   c. group of people who all want to do the same thing
   d. small hairs growing from the eyelids

25. **machete**
   a. long knife used in the tropics for cutting bushes
   b. long stick used in playing ball games
   c. food eaten for breakfast by people of Central America
   d. fault or mark on the skin

26. **mutter**
   a. do something unskilfully
   b. say something in a low, unclear voice
   c. milk product used on toast
   d. sheep

27. **palm**
   a. weapon which explodes
   b. give someone money
   c. kind of tree growing in the tropics
   d. room used as a court of law

28. **sombrero**
   a. kind of big, broad-rimmed hat
   b. kind of Mexican food
   c. alone, without friends
   d. quiet, still, emotionless

29. **pewter**
   a. empty
   b. soft to touch, like a teddy bear
   c. quiet or not very talkative
   d. a silver coloured metal

30. **plunge**
   a. not tell the truth
   b. deep, like the ocean
   c. take too much of any one thing
   d. jump into

31. rub
  a. move one thing against another
  b. stop or finish abruptly
  c. slowly move away from someone or something
  d. throw something high into the sky

32. saint
  a. holy or very good person
  b. sweet or sugary
  c. deep valley between tall mountains
  d. sudden attack of fainting

33. sandals
  a. kind of shoe with straps on top
  b. long leather straps used to control horses
  c. holes used by soldiers to hide from their enemies
  d. frame for holding things

34. saw
  a. tool used for getting water from a stream
  b. kind of tool with toothed edge used for cutting wood
  c. kind of utensil used for eating
  d. kill quickly and without mercy

35. scatter
  a. throw here and there
  b. pick up things from various places
  c. plate used for serving food
  d. break something (like glass)

36. scorpion
  a. small animal with a sting in its tail
  b. small, soft, friendly animal kept as a pet in some countries
  c. kind of ship used in the navy
  d. result at the end of a football match

37. shrug
  a. raise your shoulders
  b. stop without reason
  c. small hole made by a golf club
  d. dirty or untidy state

38. shuffle
  a. get into a fist fight
  b. kind of card game
  c. walk without lifting your feet
  d. clearly make known

39. sidewalk
  a. dance step
  b. hop up and down on one foot
  c. place to walk on the side of the street
  d. trail through the mountains

40. spray
  a. edge of a piece of cloth
  b. fail to take advantage of
  c. spread a liquid in very small drops
  d. swim quickly

41. streetlights
   a. portable lights used by policemen
   b. lights used to make a street brighter
   c. shops that sell lights
   d. lights used in a parade at night

42. stumble
   a. break into small pieces
   b. speak angrily about someone without them knowing
   c. complain
   d. put your foot down wrongly and almost fall over

43. sweat
   a. hit something with your hand or with some kind of tool
   b. opposite of sour
   c. liquid which comes through the skin when you are hot or frightened
   d. painful sickness affecting older people

44. taciturn
   a. take a turn (as in a card game)
   b. turn to the right
   c. not knowing where to go
   d. saying little, silent by habit

45. tamarind
   a. kind of ball game
   b. close friend
   c. name of a kind of dog
   d. tropical evergreen tree

46. universe
   a. first verse of a poem or song
   b. everything that exists
   c. lifetime work of a great writer
   d. kind of message

47. whim
   a. sudden idea or wish
   b. sound an animal makes when it is frightened
   c. kind of joke
   d. kind of ghost story told by children

48. wicker
   a. outgoing, friendly, likes talking to people
   b. play made in the game of cricket
   c. small insect that bites, like a mosquito
   d. sticks crossed over each other like in a basket

Appendix 2.4 : Questionnaire in Japanese

名前＿＿＿＿＿＿＿＿＿＿＿＿＿　（さしつかえなければ記入してください。）

各設問において、1～5のうち該当するものに　×をつけてください。

リーディング・エクササイズにヘッドフォンを使用しましたか。　はい ☐　いいえ ☐

	5	4	3	2	1
以前のコンピューターの使用頻度はどうでしたか。	大変頻繁に使用した (一週間に4～5回) ☐	割合よく使用した (一週間に1～2回) ☐	たまに使用した (一ケ月に1回) ☐	あまり使用しなかった (一年に1回) ☐	全く使用しなかった ☐
以前のマウスの使用頻度はどうでしたか。	大変頻繁に使用した (一週間に4～5回) ☐	割合よく使用した (一週間に1～2回) ☐	たまに使用した (一ケ月に1回) ☐	あまり使用しなかった (一年に1回) ☐	全く使用しなかった ☐
マウスを使っての指示はむずかしい。	全く同感である ☐	ある程度そう思う ☐	どちらともいえない ☐	そうは思わない ☐	全く反対意見である ☐
自分はコンピューターが好きだ。	全く同感である ☐	ある程度そう思う ☐	どちらともいえない ☐	そうは思わない ☐	全く反対意見である ☐
このストーリーには、簡単な単語が使われていた。	全く同感である ☐	ある程度そう思う ☐	どちらともいえない ☐	そうは思わない ☐	全く反対意見である ☐
このストーリーの内容（筋書き）は簡単に理解できた。	全く同感である ☐	ある程度そう思う ☐	どちらともいえない ☐	そうは思わない ☐	全く反対意見である ☐
おもしろいストーリーだった。	全く同感である ☐	ある程度そう思う ☐	どちらともいえない ☐	そうは思わない ☐	全く反対意見である ☐

自分はリーディング中にストーリーを聴くのが好きだ。	全く同感である ☐	ある程度そう思う ☐	どちらともいえない ☐	そうは思わない ☐	全く反対意見である ☐
ストーリーを読むのにコンピューターを使うのは楽しかった。	全く同感である ☐	ある程度そう思う ☐	どちらともいえない ☐	そうは思わない ☐	全く反対意見である ☐
自分は普通の本を使ってストーリーを読む方が好きだ。	全く同感である ☐	ある程度そう思う ☐	どちらともいえない ☐	そうは思わない ☐	全く反対意見である ☐
コンピューターを使ったリーディングを、またやってみたい。	全く同感である ☐	ある程度そう思う ☐	どちらともいえない ☐	そうは思わない ☐	全く反対意見である ☐
コンピューターの使用は言語習得に効果的である。	全く同感である ☐	ある程度そう思う ☐	どちらともいえない ☐	そうは思わない ☐	全く反対意見である ☐
もっと練習する機会があれば、コンピューターを使ったリーディングをより一層楽しめると思う。	全く同感である ☐	ある程度そう思う ☐	どちらともいえない ☐	そうは思わない ☐	全く反対意見である ☐
コンピューターを使ったリーディングは時間の無駄である。	全く同感である ☐	ある程度そう思う ☐	どちらともいえない ☐	そうは思わない ☐	全く反対意見である ☐
生徒全員にむかって、先生が声に出して読むのがいい。	全く同感である ☐	ある程度そう思う ☐	どちらともいえない ☐	そうは思わない ☐	全く反対意見である ☐
英語でのリーディングはむずかしい。	全く同感である ☐	ある程度そう思う ☐	どちらともいえない ☐	そうは思わない ☐	全く反対意見である ☐
英語のショート・ストーリーが好きだ。	全く同感である ☐	ある程度そう思う ☐	どちらともいえない ☐	そうは思わない ☐	全く反対意見である ☐
日本語での読書が好きだ。	全く同感である ☐	ある程度そう思う ☐	どちらともいえない ☐	そうは思わない ☐	全く反対意見である ☐
ストーリーはどのくらい理解できましたか。	0-10% ☐	10-25% ☐	25-50% ☐	50-75% ☐	75-100% ☐
単語の意味はどのくらい理解できましたか。	0-10% ☐	10-25% ☐	25-50% ☐	50-75% ☐	75-100% ☐

下記の質問におこたえ下さい。

1．学習者の立場から見て、コンピューターでのリーディングをもっと簡単にするためには、どうすればよいと思いますか。

2．コンピューターによるリーディングをもっと面白くするには、どんなものを取り入れるとよいと思いますか。

3．ご意見、ご質問など、なんでもお書き下さい。

ありがとうございました。

Appendix 2 .5 : Questionnaire in English

	5	4	3	2	1
Item 1 *How much have you used computers before?*					
Item 2 *How much have you used a mouse before?*					
Item 3 *Pointing with the mouse was difficult.*					
Item 4 *I like computers.*					
Item 5 *The words in the story were easy*					
Item 6 The content (plot) of the story was easy					
Item 7 *The story was interesting.*					
Item 8 *I like listening to stories while reading.*					
Item 9 *I enjoyed using the computer to read the story.*					
Item 10 *I would prefer to read the story in a normal book.*					
Item 11 *I would like to try reading on the computer again.*					
Item 12 *Computers are good for learning languages.*					
Item 13 *I would enjoy reading on the computer more if I could get more practice in doing it.*					
Item 14 *Using the computers for reading is a waste of time.*					
Item 15 *I like reading while listening with a teacher.*					
Item 16 *Reading in English is difficult.*					
Item 17 *I like reading short stories in English.*					
Item 18 *I like reading in my own language.*					
Item 19 *How much of the content of the story did you understand?*					
Item 20 *How much vocabulary did you understand?*					

Appendix 3

Appendix 3 .1 : Statistical Formulae used for Data Analysis

Statistical packages use a variety of names for their statistical procedures. To be clear which procedure was used, the following section lists the test used in Chapter 4 and gives the formula for each.

SPSS and the Chi-Square Distribution

The χ^2 test was performed using SPSS on the frequency results. SPSS computes a one sample χ^2 test with the following formula:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Where O_i is the observed frequency for the *i*th category, E_i is the expected frequency for the *i*th category, and *k* is the number of categories (Noru_is, M.J., SPSS Ltd.1993). The observed frequency is the actual count or respondent's answers and the expected frequency is the frequency in the cell that would be expected if the null hypothesis were true (i.e., the sample size *N* divided by the number of categories *i*).

For the χ^2 procedure to test hypotheses that row and column variables are independent, SPSS uses the Pearson χ^2. The formula is as follows:

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Where *i* is the row and *j* is the column and therefore O_{ij} is the observed frequency of cell *ij* and E_{ij} is the the expected frequency of cell *ij*.

SPSS and the t-test Statistic

SPSS calculates the t value using the following formula:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}}$$

Where $\overline{X}_1$ is the sample mean of group 1, S_1^2 is the variance and N_1 is the sample size (Noru_is and SPSS, 1993).

SPSS and the Pearson Correlation

SPSS calculates the Pearson correlation coefficient (denoted by r) as:

$$r = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{(N-1)S_X S_Y}$$

where N is the number of cases and S_X and S_Y are the standard deviations of the two variables (Noru_is and SPSS, 1993).

SPSS and the Spearman Correlation

SPSS calculates the Spearman Correlation (denoted as ρ) as:

$$\rho = 1 - \frac{6(\sum d^2)}{N(N^2 - 1)}$$

Appendix 3 .2 : Post-Test Raw Scores

E-Group Score	E-group Stdnt	C-Group Score	C-Group Stdnt
14	26	10	46
15	17	12	35
16	9	14	49
18	6	18	40
20	5	21	54
20	18	22	57
21	13	23	55
22	7	23	45
22	24	24	51
22	22	26	47
23	19	26	38
23	14	27	52
24	11	27	37
25	16	27	41
25	4	27	39
26	25	28	43
26	29	28	42
27	23	28	58
27	28	29	56
27	21	30	53
27	12	30	32
28	10	31	59
30	8	31	48
31	20	32	61
32	15	33	33
35	3	35	34
36	2	35	31
37	1	36	50
39	30	36	44
		38	36

Appendix 3 .3 : Click Data

Student Number	Score	Group	Total Clicks	Total Sentences	Repeats of Sentences	Net of Sentences	Total Words	Repeats of Words	Net of Words
3	27	E	26	17	0	17	9	0	9
4	20	E	108	108	17	91	0	0	0
5	15	E	204	198	110	88	6	2	4
7	14	E	130	126	37	89	4	2	2
8	27	E	94	94	6	88	0	0	0
12		E	46	35	0	35	11	4	7
13	31	E	296	292	195	97	4	0	4
15	22	E	109	108	16	92	1	0	1
16	26	E	126	124	34	90	2	0	2
36	32	E	194	188		91	6	1	5
37	27	E	202	172	85	87	30	13	17
41	25	E	254	251	160	91	3	0	3
44	37	E	170	168	73	95	2	1	1
48	35	E	116	110	32	78	6	2	4
49	25	E	23	23	8	15	0	0	0
53	39	E	104	104	19	85	0	0	0
55	36	E	234	232	138	94	2	1	1
56	22	E							
68	30	E	150	141	51	90	9	3	6
59	16	E							
60	28	E	98	94	11	83	4	0	4
62	24	E	139	135	46	89	4	1	3
63	27	E	105	104	13	91	1	0	1
64	21	E	22	7	1	6	15	3	12
65	23	E	133	130	58	72	3	1	2
71	18	E	82	78	5	73	4	1	3
72	20	E							
78	28	E	100	100	11	89	0	0	0
79	30	E	101	101	12	89	0	0	0

endix 3 .4 : Questionnaire Frequency Data

Key
Very Often = VO
Often = O
Once in a while = OW
Not very often = NVO
Never = Nill

	Control Group							Experimental Group						
	Number in each category							Number in each category						
	VO	O	OW	NVO	Nill	*N*		VO	O	OW	NVO	Nill	*N*	Avg.
Question 1 *How much have you used computers before?*	2	3	3	5	7	20		1	6	1	9	6	23	
Question 2 *How much have you used a mouse before*	0	2	1	8	9	20		1	1	1	6	14	23	

Key:
Strongly Agree = SA
Agree = A
Don't Know = ?
Disagree = D
Strongly Disagree = SD

	Control Group							Experimental Group						
	Number in each category							Number in each category						
	SA	A	?	D	SD	*N*	Avg.	SA	A	?	D	SD	range	Avg.
Question 3 *Pointing with the mouse was difficult.*	1	4	1	14	0	20		0	8	5	8	2	23	
Question 4 *I like computers.*	0	9	5	6	0	20		2	8	8	4	1	23	
Question 5 *The words in the story were easy*	0	5	6	8	1	20		1	3	4	15	0	23	
Question 6 The content (plot) of the story was easy	2	8	5	5	0	20		1	6	7	9	0	23	

Question 7 *The story was interesting.*	2	8	8	2	0	20		1	8	5	9	0	23	
Question 8 *I like listening to stories while reading.*	4	4	6	4	2	20		5	9	3	6	0	23	
Question 9 *I enjoyed using the computer to read the story.*	3	7	8	1	1	20		7	12	1	3	0	23	
Question 10 *I would prefer to read the story in a normal book.*	3	5	7	5	0	20		4	8	9	2	0	23	
Question 11 *I would like to try reading on the computer again.*	7	10	2	1	0	20		8	11	1	3	0	23	
Question 12 *Computers are good for learning languages.*	3	11	5	1	0	20		10	10	3	0	0	23	
Question 13 *I would enjoy reading on the computer more if I could get more practice in doing it.*	10	6	2	1	1	20		14	8	0	1	0	23	
Question 14 *Using the computers for reading is a waste of time.*	0	1	1	17	1	20		0	0	1	16	6	23	
Question 15 *I like reading while listening with a teacher.*	2	3	7	6	2	20		2	5	7	7	2	23	
Question 16 *Reading in English is difficult.*	3	9	4	4	0	20		4	12	6	1	0	23	
Question 17 *I like reading short stories in English.*	5	11	4	0	0	20		8	8	6	1	0	23	
Question 18 *I like reading in my own language.*	7	7	4	2	0	20		8	12	3	0	0	22	

	Control Group							Experimental Group						
	Number in each category							Number in each category						
	75-100%	50-75%	25-50%	10-	0-10%			75-100%	50-75%	25-50%	10-25%	0-10%		
Question 19 *How much of the content of the story did you understand?*	8	6	3	3	0	20		6	3	10	3	1	23	
Question 20 *How much vocabulary did you understand?*	2	10	4	4	0	20		1	9	9	4	0	23	

Appendix 3 .5 : Questionnaire Responses

note: questions 19 & 20 are opposite from the way they appear on the actual questionnaire (e.g., 5 is 75-100% and 1 is 0-10% in this table)

group	item / Stud	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
C	22	1	1	4	2	2	3	5	2	4	2	4	4	5	2	3	2	4	2	5	4
C	19	1	1	2	4	3	3	3	4	5	3	5	4	5	2	2	3	4	4	4	3
C	21	3	3	2	4	3	5	4	4	3	2	4	4	5	2	4	4	4	3	5	4
C	43	2	2	2	4	2	2	3	4	3	4	5	5	5	1	4	4	5	4	2	2
C	23	4	4	2	2	2	2	4	1	3	3	5	4	5	2	3	3	4	5	4	3
C	24	2	2	4	3	3	3	3	2	4	5	4	4	4	2	3	4	4	4	4	4
C	25	2	2	3	2	3	2	3	5	4	2	5	5	5	2	3	5	4	3	2	2
C	26	5	2	2	2	2	2	2	1	1	5	3	4	1	2	3	5	3	5	3	3
C	27	5	2	2	2	3	4	3	3	3	4	4	4	4	2	1	4	3	5	5	4
C	28	3	2	4	4	4	4	4	4	4	4	4	4	4	2	2	2	4	4	5	5
C	29	1	1	2	4	4	4	4	3	4	3	5	4	4	2	4	4	5	3	4	4
C	30	2	1	2	4	4	4	4	3	3	2	4	4	5	2	2	4	4	2	5	4
C	32	1	1	2	3	2	3	3	3	3	3	4	4	5	2	3	4		4	3	2
C	33	4	4	2	4	2	4	4	3	5	4	5	3	5	2	2	4	4	4	4	4
C	34	1	1	4	3	2	4	4	2	4	3	5	3	5	2	2	3	4	5	4	4
C	35	2	2	2	3	2	4	2	5	2	3	3	2	3	2	1	4	5	5	5	4
C	36	4	1	5	2	3	3	3	2	4	3	4	3	4	3	3	3	3	5	3	3
C	37	1	1	2	4	4	4	5	5	5	4	4	3	4	2	5	2	5	3	5	4
C	38	1	1	2	3	1	2	3	3	3	2	4	5	3	2	2	2	3	4	2	2
C	20	3	2	2	4	4	5	4	5	3	5	2	3	2	4	5	5	5	5	5	5
E	3	1	1	4	3	2	4	2	2	5	4	5	4	5	1	2	4	3	4	5	4
E	10	2	2	4	4	2	2	3	2	4	4	4	4	5	2	3	4	3	5	3	3
E	9	4	1	1	4	2	3	4	3	4	3	4	4	5	1	1	4	3	5	2	2
E	8	2	1	3	3	3	4	4	2	4	3	4	3	4	2	3	4	4	4	3	3
E	7	2	2	3	1	3	3	4	4	4	4	4	4	5	1	2	4	4	4	3	3
E	6	3	1	4	4	4	4	3	4	5	2	5	5	5	2	2	4	5	5	5	4
E	11	1	1	2	3	2	2	2	4	4	4	4	3	4	2	4	5	5	4	4	4
E	4	2	2	2	4	2	3	4	4	5	3	5	4	5	2	3	4	4	3	4	4
E	16	4	4	2	4	3	5	5	2	4	3	3	4	4	2	4	3	2	3	5	5
E	2	2	2	4	2	2	2	4	4	2	4	2	4	4	2	2	3	4	4	3	3
E	5	2	1	4	4	2	2	3	3	4	4	4	4	5	2	3	3	3	4	2	3
E	12	1	1	3	3	2	2	2	4	4	3	4	5	4	2	3	4	3	4	3	3
E	13	2	1	4	2	2	3	2	5	2	5	2	5	2	2	2	5	5	5	5	4
E	15	2	2	2	3	2	4	4	4	5	3	4	4	5	2	4	4	4	5	5	4
E	31	4	3	2	2	5	4	4	2	3	5	5	5	5	2	5	4	5	5	3	3
E	17	2	2	4	2	4	3	3	4	2	5	2	3	4	3	3	3	5	5	1	2
E	18	5	5	2	3	2	4	2	5	4	3	4	5	4	2	4	3	4	3	5	4
E	42	1	1	2	3	2	2	2	5	5	4	5	4	5	2	4	5	5	5	3	3
E	1	1	1	3	4	2	2	2	4	4	2	4	5	4	2	2	3	4	4	2	2
E	41	4	1	2	3	4	3	2	5	5	3	5	5	5	1	3	2	3	4	3	4
E	40	4	1	3	5	3	2	3	5	5	5	5	5	5	2	5	4	5	4	3	2
E	39	4	1	4	5	2	2	4	2	4	4	5	5	5	1	1	4	5	4	3	3
E	14	1	1	1	4	2	3	2	3	4	3	4	5	5	1	2	5	4	4	4	4

Appendix 3 .6 : Uncollapsed Data for Items 19 and 20

Item 19

The following table shows the uncollapsed data from item 19:

Association Between Sound and Overall Comprehension Using all Likert Scale Points

	Comprehension										
	0-10%		10-25%		25-50%		50-75%		75-100%		Row Total
	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs
Without Sound	0	.5	3	2.8	3	6.0	6	4.2	8	6.5	20
With Sound	1	.5	3	3.2	10	7.0	3	4.8	6	7.5	23
Column Total	1		6		13		9		14		43

The observed frequencies are indicated by 'Obs' and the expected frequencies are indicated by 'Exp'.

Item 20

The following table shows the uncollapsed data from item 20:
As with item 19, the next table shows the frequency responses for all the points on the 5-point Likert scale of item 20:

Frequency of Various Responses to Item 20 Using all Likert Scale Points

	Comprehension										
	0-10%		10-25%		25-50%		50-75%		75-100%		Row Total
	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs
Without Sound	0		4	3.7	4	6.0	10	8.8	2	1.4	20
With Sound	0		4	4.3	9	7.0	9	10.2	1	1.6	23
Column Total	0		8		13		19		3		43

Appendix 3.7 : Details of Item 10 Chi-Square Distribution

χ^2 Test over all Categories:

Category	Observed	Expected	Residual
strongly disagree	0	8.6	-8.6
disagree	7	8.6	-1.6
don't know	16	8.6	7.4
agree	13	8.6	4.4
strongly agree	7	8.6	-1.6
Total *N*	43		
χ^2	17.8140		
DF	4		
Significance	.0013		

Item 10 with agree *Categories Combined*

category 1 = *agree* and *disagree*
category 2 = all other categories

Category	Cases Observed	Expected	Residual
1.00	20	21.50	-1.50
2.00	23	21.50	1.50
	--		
Total	43		

Chi-Square	D.F.	Significance
.2093	1	.6473

Item 10 with Agree *and* Don't Know *Categories Combined*

category 1 = *agree* and *dont' know*
category 2 = both *disagree* categories

Category	Cases Observed	Expected	Residual
1.00	36	21.50	14.50
2.00	7	21.50	-14.50
	--		
Total	43		

Chi-Square	D.F.	Significance
19.5581	1	.0000

Item 10 with **Don't Know** *Category*

Category 1 = *don't know*
Category 2 = All other categories

Category	Cases Observed	Expected	Residual
1.00	16	21.50	-5.50
2.00	27	21.50	5.50
	--		
Total	43		

Chi-Square	D.F.	Significance
2.8140	1	.0934

Item 10 with **don't know** *and combined* **agree** *Categories*

Category	Cases Observed	Expected	Residual
1.00	16	18.00	-2.00
2.00	20	18.00	2.00
	--		
Total	36		

Chi-Square	D.F.	Significance
.4444	1	.5050

Item 10 with **don't know** *and combined* **disagree** *Categories*

Category	Cases Observed	Expected	Residual
1.00	16	11.50	4.50
2.00	7	11.50	-4.50
	--		
Total	23		

Chi-Square	D.F.	Significance
3.5217	1	.0606

Appendix 3 .8 : Details of Item 1 Chi-Square Distribution

Item 1 Computer Experience

Category 1 = *experienced*
Category 2 = *inexperienced*

Category	Cases Observed	Expected	Residual
1.00	12	21.50	-9.50
2.00	31	21.50	9.50
	--		
Total	43		

Chi-Square	D.F.	Significance
8.40	1	.0038

Item 2: Mouse Experience

Category 1 = *experienced*
Category 2 = *inexperienced*

Category	Observed	Expected	Residual
1.00	4	21.50	-17.50
2.00	39	21.50	17.50
	--		
Total	43		

Chi-Square	D.F.	Significance
28.49	1	.0000

Appendix 3 .9 : Details of Items 5 and 6 Chi-Square Distribution

Chi-Square Test Across all Categories, Both Groups Combined

Item 5

Category	Cases Observed	Expected	Residual
1.00	1		-7.60
2.00	23	8.60	14.40
3.00	10	8.60	1.40
4.00	8	8.60	-.60
5.00	1	8.60	-7.60
Total	43		

Chi-Square	D.F.	Significance

37.8140	4	.0000

Chi-Square Test Across all Categories, both Groups Combined

Item 6

Category	Cases Observed	Expected	Residual	
1.00	0	8.60	-8.60	
2.00	14	8.60	5.40	
3.00	12	8.60	3.40	
4.00	14	8.60	5.40	
5.00	3	8.60	-5.60	
Total	43			
Chi-Square		D.F.		Significance
20.3721		4		.0004

<u>Appendix 3 .10 : Item 5</u>

Chi-Square Test: Combined Agree against All Others

1 = Agree
2 = All Others

Category	Cases Observed	Expected	Residual	
1.00	9	21.50	-12.50	
2.00	34	21.50	12.50	
	43			
Chi-Square		D.F.		Significance
14.5349		1		.0001

Chi-Square Test Combined Don't Know and Agree against Disagree

1 = Combined
2 = Disagree

Category	Cases Observed	Expected	Residual	
1.00	19	21.50	-2.50	
2.00	24	21.50	2.50	
Total	43			
Chi-Square		D.F.		Significance
.5814		1		.4458

Chi-Square Test: Don't Know against All Others

1 = Don't know
2 = All Others

Category	Cases Observed	Expected	Residual
1.00	10	21.50	-11.50
2.00	33	21.50	11.50

Total	43			
Chi-Square		D.F.		Significance
12.3023		1		.0005

Chi-Square Test: Don't Know against Combined Agree

1 = Don't Know
2 = Combined Agree

Category	Cases Observed	Expected	Residual	
1.00	10	9.50	.50	
2.00	9	9.50	-.50	
Total	19			
Chi-Square		D.F.		Significance
.0526		1		.8185

Chi-Square Test: Don't Know against Disagree

1 = Don't Know
2 = Disagree

Category	Cases Observed	Expected	Residual	
1.00	10	17.00	-7.00	
2.00	24	17.00	7.00	
Total	34			
Chi-Square		D.F.		Significance
5.7647		1		.0164

Appendix 3 .11 : Item 6

Chi-Square Test: Combined Agree against All Others

1 = Agree
2 = All Others

Category	Cases Observed	Expected	Residual	
1.00	17	21.50	-4.50	
2.00	26	21.50	4.50	
Total	43			
Chi-Square		D.F.		Significance
1.8837		1		.1699

Chi-Square Test: Combined Disagree against All Others

1 = All Others
2 = Disagree

Category	Cases Observed	Expected	Residual
1.00	29		7.50
2.00	14	21.50	-7.50

Total	43			
Chi-Square		D.F.		Significance
5.2326		1		.0222

Chi-Square Test: Don't Know against All Others

1 = Don't Know
2 = All Others

Category	Cases Observed	Expected	Residual	
1.00	12	21.50	-9.50	
2.00	31	21.50	9.50	
Total	43			
Chi-Square		D.F.		Significance
8.3953		1		.0038

Chi-Square Test: Don't Know against Combined Agree

1 = Don't Know
2 = All Others

Category	Cases Observed	Expected	Residual	
1.00	12	14.50	-2.50	
2.00	17	14.50	2.50	
Total	29			
Chi-Square		D.F.		Significance
.8621		1		.3532

Chi-Square Test: Don't Know against Combined Disagree

1 = Don't Know
2 = Combined Disagree

Category	Cases Observed	Expected	Residual	
1.00	12	13.00	-1.00	
2.00	14	13.00	1.00	
Total	26			
Chi-Square		D.F.		Significance
.1538		1		.6949

Appendix 4

Appendix 4 .1 : Samples of Log Data Categorisations

Action	Comment
went ahead one page at:	
went back one page at:	
woke	
Iwoke	
Hotsteam	
AGrey	
IJumped	
Iwent	
went ahead one page at:	skipped through pages without listening
went ahead one page at:	
went ahead one page at:	
went ahead one page at:	
went ahead one page at:	
AllRight	
went ahead one page at:	
went back one page at:	
went back one page at:	
went back one page at:	
went back one page at:	
went back one page at:	
went back one page at:	
went ahead one page at:	
went ahead one page at:	
tamarinds	
hummed	
TheNight	sporadic listening
went ahead one page at:	
Don'tMove	
went ahead one page at:	
went ahead one page at:	
TheFlame	
went ahead one page at:	
went back one page at:	
went back one page at:	
went back one page at:	
went back one page at:	
went back one page at:	
went back one page at:	
went back one page at:	
went ahead one page at:	
went ahead one page at:	
went back one page at:	
hammock	
scorpion	
enormous	
went ahead one page at:	
fumbled	
cobblestone	
IFumbled	
went ahead one page at:	
WindWhistled	
bivouacked	
blink	
MyAction	
scattered	
WhatWord	
IThrew	
comet	
went ahead one page at:	
went ahead one page at:	
went ahead one page at:	
went ahead one page at:	
went ahead one page at:	
went ahead one page at:	
went ahead one page at:	
THIS CONCLUDES LOG OF: aiko	No pattern emerging: Category 4
HeWasSmall	started listening in the middle of the story
HeWasSmall	
went ahead one page at:	
InHisRight	
InHisRight	
LetMe	
HeStruck	
GrabbingMy	
IKnelt	
WithOne	
HeBent	
WithOne	
HeBent	
IClosed	
went ahead one page at:	listened sporadically
IPulled	
StumblingFalling	
IRan	
WhenI	
IWentIn	
TheNext	
StumblingFalling	
IRan	
WhenI	
IWentIn	
went ahead one page at:	
went back one page at:	went back over the pages but did not listen: category 4

went back one page at:	
went back one page at:	
went back one page at:	
went back one page at:	
went back one page at:	
went back one page at:	
went ahead one page at:	THIS CONCLUDES LOG OF: akiko-i.txt mine edited
woke	
woke	
woke	
Iwoke	
Hotsteam	
Hotsteam	
grey-wing	
AGrey	more than 50% listened to
IJumped	
IJumped	
IWent	
OneCould	
IRubbed	
IRan	
went ahead one page	
AtTheDoor	
SittingOn	
InAHoarse	
WhereAre	
ToTake	
AndNot	
IShrugged	
AtFirst	more than 50% listened to
IFumbled	
ILit	
SuddenlyThe	
IStopped	
WindWhistled	
went ahead one page	
IBreathed	
TheNight	
CricketsBivouacked	
IRaised	
Upthere	
IThought	
MyAction	
WhatWord	
WhoSpeaks	
ToWhom	more than 50% listened to
IThrew	
FallingIt	
IWalked	
IFelt	
IFelt	
TheNight	
AsICrossed	
ITurned	
went ahead one page	
IHurried	
AFew	
AFew	
IDidn't	
IDidn't	more than 50% listened to
ITried	
ICouldn't	
SuddenlyIStopped	
BeforeI	
Don'tMove	
Don'tMove	
WithoutTurning	
WithoutTurning	
YourEyes	
MyEyes	
WhatDo	
LookI've	
I'llGive	
went ahead one page	
Don'tBe	
I'mOnly	
ButWhy	
MyGirlfriend	
SheWants	
AndAround	more than 50% listened to
MyEyes	
They'reBrown	
Don'tTry	
IKnow	
Don'tTake	
I'llGive	
Don'tPlay	
TurnAround	
TurnAround	
Don'tPlay	
HeWasSmall	
HisPalm	
HisPalm	
HisPalm	
HisPalm	
HisPalm	

HisPalm	
LetMe	
went ahead one page	
HeStruck	
GrabbingMy	
WithOne	
HeBent	more than 50% listened to
KeepThem	
TheFlame	
ILeaned	
went ahead one page	
IPulled	
StumblingFalling	
IRan	
IRan	
WhenI	
IWentIn	
TheNext	
went back one page	
went back one page	
went back one page	
went back one page	
went back one page	
went back one page	
went back one page	
woke	started again at the beginning
Iwoke	
Hotsteam	
AGrey	
IJumped	
IWent	
OneCould	read here and there on the second pass
IReturned	
IRubbed	
IRubbed	
IRan	
AtTheDoor	
SittingOn	
InAHoarse	THIS CONCLUDES LOG OF: akiko-y.txt
Iwoke	
Hotsteam	
AGrey	
IJumped	
IWent	
IWent	
OneCould	50% listened to on page
IReturned	
IRubbed	
Iwoke	
went ahead one page at:	
IRan	
AtTheDoor	
SittingOn	
InAHoarse	
WhereAre	
ToTake	
AndNot	50% listened to on page
You'dBetter	
shrugged	
IShrugged	
AtFirst	
IFumbled	
ILit	
SuddenlyThe	
IStopped	
IStopped	
went ahead one page at:	
WindWhistled	
IBreathed	
TheNight	
CricketsBivouacked	
IRaised	
Upthere	
IThought	50% listened to on page
MyAction	
WhatWord	
WhoSpeaks	
ToWhom	
IThrew	
FallingIt	
IWalked	
IFelt	
TheNight	
AsICrossed	
went ahead one page at:	
ITurned	
IHurried	
AFew	
IDidn't	
ITried	
ICouldn't	
SuddenlyIStopped	50% listened to on page
BeforeI	
Don'tMove	
WithoutTurning	
YourEyes	
MyEyes	
WhatDo	
LookI've	
went ahead one page at:	
NotMuch	
I'llGive	
Don'tKill	
Don'tBe	
I'mOnly	
ButWhy	
MyGirlfriend	50% listened to on page
SheWants	
AndAround	
MyEyes	

They'reBrown	
Don'tTry	
IKnow	
Don'tTake	
I'llGive	
Don'tPlay	
TurnAround	
ITurned	
HeWasSmall	
went ahead one page at:	
InHisRight	
LetMe	
HeStruck	
GrabbingMy	
IKnelt	
WithOne	
HeBent	
IClosed	
KeepThem	
IOpened	50% listened to on page
TheFlame	
AllOf	
AllRight	
BeatIt	
went ahead one page at:	
HeVanished	
ILeaned	
IPulled	
StumblingFalling	
IRan	
WhenI	
IWentIn	
TheNext	read through once completely
went back one page at:	
went back one page at:	
went back one page at:	
went back one page at:	
went back one page at:	
went back one page at:	
TheBlue	
TheBlue	
TheBlue	
Iwoke	
Hotsteam	
AGrey	scanned through a second time
IJumped	
IWent	
OneCould	
IReturned	
went back one page at:	
went back one page at:	
went back one page at:	

THIS CONCLUDES LOG OF: aya.txt

{filename |AYA.TXT}

Iwoke	
Hotsteam	
AGrey	
IJumped	
IWent	
OneCould	
IReturned	
IRubbed	
went ahead one page	
IRan	
AtTheDoor	
SittingOn	
InAHoarse	
WhereAre	
ToTake	
AndNot	
You'dBetter	
shrugged	
IShrugged	
AtFirst	
IFumbled	
ILit	
SuddenlyThe	
IStopped	
went ahead one page	
went ahead one page at	went back and forth here
went back one page	
hummed	
WindWhistled	
IBreathed	
TheNight	
CricketsBivouacked	
IRaised	
Upthere	
IThought	
MyAction	
WhatWord	
WhoSpeaks	
ToWhom	
IThrew	
FallingIt	
IWalked	
IFelt	
TheNight	
AsICrossed	
went ahead one page	
ITurned	
IHurried	
AFew	
IDidn't	
ITried	
ICouldn't	
SuddenlyIStopped	

BeforeI
Don'tMove
WithoutTurning
YourEyes
MyEyes
WhatDo
LookI've
went ahead one page
NotMuch
I'llGive
Don'tKill
Don'tBe
I'mOnly
ButWhy
MyGirlfriend
SheWants
AndAround
Don'tTry
MyEyes
They'reBrown
Don'tTry
IKnow
Don'tTake
Don'tTake
I'llGive
Don'tPlay
TurnAround
ITurned
HeWasSmall
went ahead one page
InHisRight
LetMe
HeStruck
GrabbingMy
IKnelt
WithOne
HeBent
IClosed
KeepThem
IOpened
TheFlame
AllOf
AllRight
BeatIt
went ahead one page
HeVanished
ILeaned
IPulled
StumblingFalling
IRan
WhenI
IWentIn
TheNext
went ahead one page
went ahead one page at:
went back one page
went back one page
went back one page
went back one page
went back one page
went back one page
went back one page
went back one page
went back one page
went ahead one page
woke
woke
Iwoke
Hotsteam
AGrey
IJumped
IWent
OneCould
OneCould
IReturned
IRubbed
went ahead one page
IRan
AtTheDoor
SittingOn
WhereAre
ToTake
AndNot
IShrugged
AtFirst
IFumbled
ILit
SuddenlyThe
IStopped
went back one page
WindWhistled
IBreathed
TheNight
CricketsBivouacked
IRaised
Upthere
IThought
MyAction
WhatWord
WhoSpeaks
ToWhom
IThrew
FallingIt
comet
IWalked
IFelt
TheNight
AsICrossed
went ahead one page
ITurned
IHurried
AFew

IDidn't
ITried
SuddenlyIStopped
BeforeI
Don'tMove
WithoutTurning
YourEyes
MyEyes
WhatDo
LookI've
went ahead one page
NotMuch
I'llGive
Don'tKill
Don'tBe
ButWhy
MyGirlfriend
whim
SheWants
MyEyes
They'reBrown
Don'tTry
IKnow
Don'tTake
I'llGive
Don'tTake
Don'tPlay
TurnAround
HeWasSmall
HisPalm
HisPalm
HisPalm
HisPalm
HisPalm
HisPalm
HisPalm
HisPalm
went ahead one page
InHisRight
LetMe
GrabbingMy
WithOne
HeBent
KeepThem
IOpened
TheFlame
AllOf
went ahead one page
ILeaned
IPulled
StumblingFalling
IRan
WhenI
IWentIn
boardinghouse2 — Read through almost completely twice
went ahead one page
went ahead one page
went ahead one page
went ahead one page
went back one page
went back one page
went back one page
went back one page
went ahead one page
went ahead one page

went ahead one page
went ahead one page
went ahead one page
went back one page
went back one page
went back one page — THIS CONCLUDES LOG OF: emi-h.txt

woke
Iwoke
Hotsteam
AGrey
IJumped
IWent
OneCould
IReturned
IRubbed
went ahead one page at:
IRan
AtTheDoor
SittingOn
InAHoarse
WhereAre
ToTake
AndNot
You'dBetter
shrugged
IShrugged
IShrugged
AtFirst
IFumbled
ILit
SuddenlyThe
IStopped
went ahead one page at:
WindWhistled
IBreathed
TheNight
CricketsBivouacked
IRaised
Upthere
IThought
MyAction
WhatWord
WhoSpeaks
ToWhom
IThrew

FallingIt
IWalked
IFelt
TheNight
AsICrossed
went ahead one page at:
ITurned
IHurried
AFew
IDidn't
ITried
ICouldn't
SuddenlyIStopped
BeforeI
Don'tMove
WithoutTurning
YourEyes
MyEyes
WhatDo
LookI've
went ahead one page at:
NotMuch
I'llGive
Don'tKill
Don'tBe
I'mOnly
ButWhy
MyGirlfriend
SheWants
AndAround
MyEyes
They'reBrown
Don'tTry
IKnow
Don'tTake
I'llGive
Don'tPlay
TurnAround
ITurned
HeWasSmall
went ahead one page at:
InHisRight
LetMe
HeStruck
GrabbingMy
IKnelt
WithOne
HeBent
IClosed
KeepThem
IOpened
TheFlame
AllOf
AllRight
BeatIt
went ahead one page at:
HeVanished
ILeaned
IPulled
StumblingFalling
IRan
WhenI
IWentIn
TheNext
went ahead one page at:
went back one page at:
went back one page at:
went back one page at:
went back one page at:
went back one page at:
went back one page at:
went back one page at:
woke
Iwoke
Hotsteam
AGrey
IJumped
IWent
OneCould
IReturned
IRubbed
went ahead one page at:
IRan
AtTheDoor
SittingOn
WhereAre
ToTake
AndNot
You'dBetter
shrugged
shrugged
IShrugged
AtFirst
IFumbled
ILit
SuddenlyThe
IStopped
went ahead one page at:
WindWhistled
IBreathed
TheNight
CricketsBivouacked
IRaised
Upthere
IThought
MyAction
WhatWord

went back one page at:
went back one page at:
went back one page at:
went back one page at:

Read through onec plus scanning on second pass: category 2

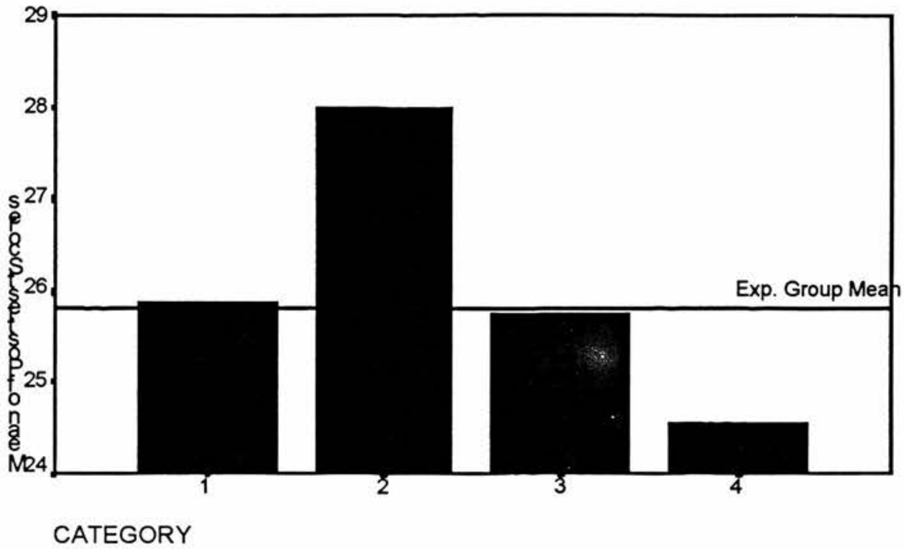
went back one page at:

THIS CONCLUDES LOG OF: Hiromi-m.txt

(filename |HIROMI-M.TXT)

Appendix 4 .2 : Details of Analysis of Variance Data for Log Categorisations

Performance of Various Groups on the Post-test: Category v Post-test Mean



Means of Categories 1 through 4

Total Population

mean	25.77
N	(26)

Category Type

	1	2	3	4
mean	25.86	28.00	25.75	23.75
N	(7)	(7)	(4)	(8)

Analysis of Variance

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Main Effects	67.508	3	22.503	.479	.700
LOGTYPE	67.508	3	22.503	.479	.700
Explained	67.508	3	22.503	.479	.700
Residual	1033.107	22	46.959		
Total	1100.615	25	44.025		

Appendix 5

Appendix 5 .1 : Data from the Open-ended Questionnaire Items.

Note: the answers are translations from Japanese except where noted.

Question 1: What do you think could be done to make the reading exercise easier?
Question 2: What do you think could be done to make the reading exercise more interesting?
Question 3: Other comments.

Student ID

Student's

Group (C= Control; E= Experimental)

Q	Student: 17 F C
1	I'd like to use various symbols to mark learning points, and important/difficult parts of the text while reading. It would be helpful if important sentences were highlighted.
2	It would be more enjoyable if I could actually move pictures of the characters in the story, or add nice landscapes on the screen during or after the reading.
3	The only problem was that it made my eyes tired.

	Student: 18 F C (answered in English)
1	Using tape recorder with computers at the same time helps our reading quite a lot.
2	The stories of very popular of famous people who still live are very interesting for learners.
3	To use computers in order to teach grammar or sentence patterns is very useful

	Student: 25 F C
1	Activities using games would make the program more interesting and accessible.
2	Use of animation might help us understand the context. It would also encourage us to read on.
3	I was confused at first as I had almost no knowledge or experience of computers. However, I know I have to learn computer operation in order to cope with office automation.

	Student: 34 ? C
1	A teacher/instructor could give us instructions through headphones. We need to familiarise ourselves with the computer by using it regularly.
2	Games Illustrations Quizzes Graphics
3	

	Student: 35 M C
1	To make the program easier, you may wish to give some hints to some of the difficult words, but I would personally like it to remain as it is. The program enhances imagination which helps us in reading English stories.
2	I do not think the program should be changed in any way. Adding things (that are unnecessary) would make the screen too decorative (and possibly bad for the eyes).
3	I found the lesson really interesting and I guess it is mainly because I am interested in computers. Reading books can be boring—sometimes it makes me really sleepy, but the same thing can be made so enjoyable with the computer

	Student: 40 F C
1	
2	Illustrations
3	

	Student: 42 F C
1	Illustrations would help us understand the content of stories better.
2	
3	

	Student: 1 F C
1	I'm not familiar with computer use in general. I need to start with learning computers before I do reading activities.
2	I'd like some nice pictures
3	It makes me sleepy!! Reading (or looking at) alphabets is not very exciting, however, the idea of using computer for reading activity itself is very new to me and very enjoyable

	Student: 2 F C
1	Satisfactory as it is
2	I'd like more enjoyable texts. Something more exciting (i.e., something which makes you feel more involved with the story. I'd like easier texts.
3	First experience of this kind of reading exercise. Computer makes reading activities more interesting, however. I'm anxious about its effect on the learner's eyesight Using computer for few hours a week might be appropriate I'd also like to hear a teacher's pronunciation rather than reading on the computer alone.

	Student: 3 M C (answered in English)
1	I think nothing need do more. It's enough.
2	I think it is better to be take. Computer Grafics
3	

	Student: 4 F C (answered no. 1 in English)
1	I didn't know how to use the mouse but if I knew how to use it I could enjoy reading more. So I hope that we can use more computers.
2	More pictures!
3	It was a good experience to use the computer for reading English. I'd like to learn how to use the computer as it is crucial to have operational skills in this computer-oriented age.

	Student: 5 M C (answered partly in English)
1	Using the direction keys may help the program's accessibility for turning pages.
2	no comment
3	At times, sentences/paragraphs were divided in an unusual way at the end of a page--which made it difficult to read

	Student: 7 F C
1	I'd like some dictionary device attached to the program. For example, whenever I find a new/difficult word, I'd move the cursor there, and the meaning would be given. In this way, I'd skim the story first and then from the second reading, I could consolidate my understanding of the meaning of words, use and the content of the story.
2	I'd prefer the same program with more colours for example, if I find some words/difficult, I could highlight them with colours and move to the next. I can always come back to the problem parts easily with a colour marking system.
3	Although we missed the listening activity, I quite liked this new approach (as I'm one to jump on the bandwagon!) One thing which worries me is that we'll lose opportunities to write English by hand. I strongly think it's important to practice handwriting as well.

	Student: 8 F C (answered in English)
1	To make one page a little shorter and end the story where you want to know more i.e., to guess what happens next
2	You could have some quizzes after reading the story You can make some blanks in the story and students make a story by themselves and compare it with the original story afterwards. Students like to use computers in English class. They concentrate more and it's useful to use computer once a week or so.
3	

	Student: 9 M C
1	I find the activity makes my eyes really tired. Could you do something about that?
2	Add illustrations Add variety in the stories (not every story is necessarily interesting to everyone. We'd like to choose from a large number of them.
3	I was a bit drunk today. When I'm sober I'm sure I'd do better. Sorry, but this questionnaire might be a bit unclear--it's all because of alcohol!!

	Student: 10 F C (answered in English)
1	
2	If I could watch the actions of characters at the same time in the half of the screen, it would be more fun and would be easy to understand the situations.
3	

	Student: 11 F C
1	
2	
3	When a word is divided at the end of the page, I find I really difficult to read. Because of that I'd loose track of the story and become confused.

	Student: 12 M C
1	What we want is some sort of dictionary device. Some annotation may be really useful too. This could be combined with the dictionary device??
2	Background music Illustrations
3	I much prefer reading stories by computer program or cassette tapes than to traditional book-reading

	Student: 13 M C
1	If I can check the meanings of difficult words on the computer, it would help my reading greatly. I could perhaps have a little corner on the screen that gives meanings and notes for words I want to learn.
2	Not only stories but also games including comics and quizzes!!
3	I don't like computers myself, but I think I can operate it well. I reckon we are in the computer age but I really hate it!!

	Student: 14 F C (answered in English)
1	Perhaps there are lots of words which can't be understood. So why don't you prepare a small section which tells the meaning of the difficult words for students?
2	At first, the students read the story by themselves. Then after finishing their reading, they listen to the story (which the computer tells them) with effective sounds, etc. Students choose the stories. While they are reading, they could decide how the story goes--the ending, etc..
3	

	Student: 15 F E
1	Enlarge the size of the vocabulary.
2	add music, etc.
3	

	Student: 16 F E
1	
2	add illustrations add colour
3	

	Student: 19 M E
1	We need a dictionary system in the program!!
2	Students here are from all over the world--why don't you use stories about various countries? I'd like to see various topics, too--football, sports, history, world history...
3	

	Student: 20 M E (answered in English)
1	It is necessary to have a dictionary function to search difficult words.
2	I'd like to listen the whole story at once--not each sentence. It is good practice not only reading but also listening.
3	The mouse didn't respond so quickly. Sometimes I had to wait for listening the voice. Anyway, it was good fun to use computer for reading English

Y	Student: 21 M E
1	Can you make the sound respond a little faster?
2	More amusing topics please...sports, music, etc. are popular.
3	Occasionally, the voice suddenly cut out. Is this a technical matter? Is there any solution?

	Student: 22 M E
1	Illustrations would help us get the overall idea about the story. We could then read the story with a more vivid imagination and contextualise each sentence more.
2	English-English dictionary program games--activities in the style of games illustrations as above
3	This is my first experience with computer-supported reading in English. It was as natural as reading a book. I also liked using the mouse. Since we've used headphones, more audio-visual support might make the program even more effective.

	Student: 23 F E
1	You could either lower the level of vocabulary or use a dictionary program.
2	Audio support (music or sound effects) Games for activities
3	At first, I was confused as it was my first experience of computer-assisted reading. Although the operation itself was simple, the level of vocabulary was a little too high. Word-processing on computers is quite popular in Japan but computer-assisted reading is very rare. I would say that this approach is very good for learning both computer operation and English.

	Student: 24 F E (answered in English)
1	The computer program was good but it spoke too fast for me and so I it should speak more slowly.
2	As you should put pictures in the computer, they would help me understand easily (i.e., visual support—illustrations— would make the program more effective *--translator's note.*)
3	It was interesting, because I used a computer for the first time.

	Student: 26 F E
1	
2	The use of current topics would make it more interesting (although I know that this would be technically or practically difficult) I'd like a Japanese translation as well.
3	It was new to me and I enjoyed it very much. I hope we'll have more opportunities like this. I believe the program is really an effective way to improve my listening skills, but the only drawback is...it tires my eyes out!

	Student: 27 F E
1	Readers should be able to control the speed of the reading.
2	I'd like a more interesting story.
3	I think the combination of sight-reading and listening is a good way of improving our listening skills.

	Student: 28 F E
1	The slow response of the mouse irritated me a lot.
2	Operation of the mouse should be made easier. Topics for texts should be changed to something more interesting.
3	It is remarkably helpful that we can repeat words and sentences many times. I'd like to learn more words and expressions by repetition, and possibly, sentences as well. I thought it would be more effective to have a dictionary system in the program.

	Student: 29 F E
1	Normally, we do not have any opportunity to use computer-assisted learning of any kind, so it was an interesting discovery that we can practise reading English in this way. I wish we had more chance to use computers.
2	Illustrations
3	My proficiency level in English is below this reading activity. I'd like to try something of my level as I believe this is an effective way of learning English.

	Student: 30 F E
1	I would have thought the operation of using the mouse could have been made easier.
2	I think it is really interesting as it is.
3	This is an interesting approach but unfortunately, computers and software are often very costly. Can they be made more affordable in some way?

	Student: 31 ? E (answered in English)
1	It is to find the simpler way to use computer because it takes much time to master a computer.
2	More changing voice tone Including pictures
3	It is the first time to study by using computer So it was enjoyable for me to use it. I'm looking forward to spread the computer reading.

	Student: 32 M E (answered no. 3 in English)
1	Automatic mouse??
2	Illustrations for each sentence
3	It is easy to read books but computer is more difficult. Needs: pictures of paragraphs.

	Student: 33 F E
1	I'd like to control the speed of reading, particularly when I find some words difficulty to understand. It would be helpful if any new/difficult words were explained on the screen.
2	Illustrations Animation
3	I fear this approach may make my learning one-sided—either listening or reading. If I had more time and control over the reading speed, I would have found it more effective. I'd also like to change the topics of the text. It is very helpful to have an aid for pronunciation.

	Student: 36 F E
1	Make the story more interesting or use more popular and familiar stories instead, please!! If different stories are available according to the learner's level, the program would be more effective and useful.
2	
3	Occasionally I noticed distracting noises and other problems with the volume. You might wish to use more than one voice for dialogues.

	Student: 37 F E
1	A speed control device would be more effective for self-study. In any case, I think we just have to practise reading, using the program repeatedly.
2	More interesting stories please. It is confusing that a story is read by the same voice. Different voices would help us identify the characters. SE and BGM will make us motivated to read on.
3	Whoever the programmer is, he must be a GENIUS!!

	Student: 38 M E
1	I wish I could listen to difficult words and expressions at a much slower speed.
2	An English-English dictionary should be incorporated into the program.
3	Computer reading is much more user-friendly than traditional tapes, but it a more affordable price for these programs would be our dream. It is also perfect for self study.

	Student: 39 F E
1	Add a speed control device, particularly the speed of the voice.

2	More than one voice for different characters. SE and BGM
3	Very tiring for the eyes

	Student: 41 F E
1	A speed control device for listening activities would make the program much more user friendly, effective and interesting.
2	More variety in the level of listening activities. The instructions could be made easier to follow.
3	Thank you very much for a most interesting class. I would certainly like to try this kind of learning again.

	Student: 43 M E (answered in English)
1	Please don't use a mouse and don't divide a sentence into different pages.
2	It would be better to enable readers to hide the sentences (and have listening only?) Animation would be amusing
3	

	Student: 6 F E
1	I can't answer objectively since this is the first item I used computer for a reading activity. However, I'd thought combining this with listening to the teacher's voice may be even more effective and the headphones could be used for individual exercises.
2	I'd be happier if there were an audio summary of the story after each activity.
3	I like this kind of lesson very much and I believe this is very effective.

Appendix 6

Appendix 6 .1 : ThreeVersions of The Blue Bouquet with Target Words Gapped

The Blue Bouquet (Version 1)
by Octavio Paz

I woke covered with ______1______. Hot steam rose from the newly sprayed, red-brick pavement. A grey-winged butterfly, ______2______, circled the yellow light. I jumped from my hammock and crossed the room _____3_____, careful not to step on some scorpion leaving his hideout for a bit of fresh air. I went to the little window and inhaled the country air. One could hear the breathing of the night, feminine, enormous. I returned to the center of the room, emptied water from a jar into a pewter basin, and wet my towel. I rubbed my chest and legs with the soaked cloth, dried myself a little, and, making sure that no bugs were hidden in the fold of my clothes, got dressed. I ran down the green stairway. At the door of the boardinghouse I bumped into the owner, a one-eyed _____4_______ fellow. Sitting on a wicker stool, he smoked, his eye half closed. In a hoarse voice, he asked:

"Where are you going?"

"To take a walk. It's too hot."

"Hmmm--everything's closed. And no streetlights around here. You'd better stay put." I ____5____ my shoulders, muttered "back soon," and plunged into the darkness. At first I couldn't see anything. I fumbled along the ______6______ street. I lit a cigarette. Suddenly the moon appeared from behind a black cloud, lighting a white wall that was _____7_______ in places. I stopped, blinded by such whiteness. Wind whistled slightly. I breathed the air of the tamarinds. The night hummed, full of leaves and insects. Crickets bivouacked in the tall grass. I raised my head: up there the stars too had set up camp. I thought that the universe was a vast system of signs, a conversation between giant beings. My actions, the cricket's _____8_______, the star's blink, were nothing but pauses and syllables, scattered phrases from that dialogue. What word could it be, of which I was only a syllable? Who speaks the word? To whom is it spoken? I threw my cigarette down on the ______9______. Falling, it drew a shining curve, shooting out brief sparks like a tiny - ______10______.

I walked a long time, slowly. I felt free, secure between the lips that were at that moment speaking me with such happiness. The night was a garden of eyes. As I crossed the street, I heard someone come out of a doorway. I turned around, but could not ____11_______ anything. I hurried on. A few moments later I heard the dull shuffle of sandals on the hot stone. I didn't want to turn around, although I felt the shadow getting closer with every step. I tried to run. I couldn't. Suddenly I stopped short. Before I could defend myself, I felt the point of a knife in my back and a sweet voice;

"Don't move, mister, or I'll stick it in."

Without turning, I asked;

"What do you want?"

"Your eyes, mister," answered the soft, almost painful voice.

"My eyes? What do you want with my eyes? Look, I've got some money. Not much, but it's something. I'll give you everything I have if you let me go. Don't kill me."

"Don't be afraid, mister, I won't kill you. I'm only going to take your eyes."

"But why do you want my eyes?" I asked again.

"My girlfriend has this whim. She wants a _____12_____ of blue eyes. And around here they're hard to find."

"My eyes won't help you. They're brown, not blue."

"Don't try to fool me, mister. I know very well that yours are blue."

"Don't take the eyes of a fellow man. I'll give you something else."

"Don't play saint with me," he said harshly. "Turn around."

I turned. He was small and fragile. His palm _____13______ covered half his face. In his right hand he held a country _____14______ that shone in the moonlight.

"Let me see your face."

He struck another match, and put it near my eyes. Grabbing my sleeve, he ordered:

"Kneel down."

I knelt. With one hand he grabbed me by the hair, pulling my head back. He bent over me, curious and tense, while his machete slowly dropped until it grazed my eyelids. I closed my eyes.

"Keep them open," he ordered.

I opened my eyes. The flame burned my _____15______. All of a sudden he let me go.

"All right, they're not blue. Beat it." He vanished. I leaned against the wall, my head in my hands. I pulled myself together. _____16______, falling, trying to get up again. I ran for an hour through the deserted town. When I got to the plaza, I saw the owner of the _____17______, still sitting in the front of the door. I went in without saying a word. The next day I left town.

ver. 1

The Blue Bouquet (Version 2)

by Octavio Paz

I woke covered with sweat. Hot steam rose from the newly _____1_____, red-brick pavement. A grey-winged butterfly, dazzled, circled the yellow light. I jumped from my hammock and crossed the room barefoot, careful not to step on some _____2_____ leaving his hideout for a bit of fresh air. I went to the little window and _____3_____ the country air. One could hear the breathing of the night, feminine, enormous. I returned to the center of the room, emptied water from a jar into a ______4____ basin, and wet my towel. I rubbed my chest and legs with the soaked cloth, dried myself a little, and, making sure that no bugs were hidden in the _____5_____ of my clothes, got dressed. I ran down the green stairway. At the door of the boardinghouse I bumped into the owner, a one-eyed taciturn fellow. Sitting on a _____6_____ stool, he smoked, his eye half closed. In a ____7_____ voice, he asked:

"Where are you going?"

"To take a walk. It's too hot."

"Hmmm--everything's closed. And no _____8_____ around here. You'd better stay put." I shrugged my shoulders, muttered "back soon," and ____9______ into the darkness. At first I couldn't see anything. I fumbled along the cobblestone street. I lit a cigarette. Suddenly the moon appeared from behind a black cloud, lighting a white wall that was crumbled in places. I stopped, blinded by such whiteness. Wind whistled slightly. I breathed the air of the ____10______. The night hummed, full of leaves and insects. Crickets bivouacked in the tall grass. I raised my head: up there the stars too had set up camp. I thought that the universe was a vast system of signs, a conversation between giant beings. My actions, the cricket's saw, the star's ____11______, were nothing but pauses and syllables, scattered phrases from that dialogue. What word could it be, of which I was only a syllable? Who speaks the word? To whom is it spoken? I threw my cigarette down on the sidewalk. Falling, it drew a shining curve, shooting out brief sparks like a tiny comet.

I walked a long time, slowly. I felt free, secure between the lips that were at that moment speaking me with such happiness. The night was a garden of eyes. As I crossed the street, I heard someone come out of a doorway. I turned around, but could not distinguish anything. I hurried on. A few moments later I heard the dull ____12______ of sandals on the hot stone. I didn't want to turn around, although I felt the shadow getting closer with every step. I tried to run. I couldn't. Suddenly I stopped short. Before I could defend myself, I felt the point of a knife in my back and a sweet voice;

"Don't move, mister, or I'll stick it in."

Without turning, I asked;

"What do you want?"

"Your eyes, mister," answered the soft, almost painful voice.

"My eyes? What do you want with my eyes? Look, I've got some money. Not much, but it's something. I'll give you everything I have if you let me go. Don't kill me."

"Don't be afraid, mister, I won't kill you. I'm only going to take your eyes."

"But why do you want my eyes?" I asked again.

"My girlfriend has this whim. She wants a bouquet of blue eyes. And around here they're hard to find."

"My eyes won't help you. They're brown, not blue."

"Don't try to fool me, mister. I know very well that yours are blue."

"Don't take the eyes of a fellow man. I'll give you something else."

"Don't play ____13______ with me," he said harshly. "Turn around."

I turned. He was small and fragile. His ____14______ sombrero covered half his face. In his right hand he held a country machete that shone in the moonlight.

"Let me see your face."

He struck another match, and put it near my eyes. Grabbing my sleeve, he ordered:

"Kneel down."

I knelt. With one hand he grabbed me by the hair, pulling my head back. He bent over me, curious and tense, while his machete slowly dropped until it ____15______ my eyelids. I closed my eyes.

"Keep them open," he ordered.

I opened my eyes. The flame burned my lashes. All of a sudden he let me go.

"All right, they're not blue. Beat it." He vanished. I leaned against the wall, my head in my hands. I pulled myself together. Stumbling, falling, trying to get up again. I ran for an hour through the deserted town. When I got to the plaza, I saw the owner of the boardinghouse, still sitting in the front of the door. I went in without saying a word. The next day I left town.

ver. 2

The Blue Bouquet (Version 3)
by Octavio Paz

I woke covered with sweat. Hot steam rose from the newly sprayed, red-brick pavement. A ____1______ butterfly, dazzled, circled the yellow light. I jumped from my ____2______ and crossed the room barefoot, careful not to step on some scorpion leaving his ____3______ for a bit of fresh air. I went to the little window and inhaled the country air. One could hear the breathing of the night, feminine, ____4______. I returned to the center of the room, emptied water from a jar into a pewter basin, and wet my towel. I _____5_____ my chest and legs with the soaked cloth, dried myself a little, and, making sure that no bugs were hidden in the fold of my clothes, got dressed. I ran down the green stairway. At the door of the boardinghouse I bumped into the owner, a one-eyed taciturn fellow. Sitting on a wicker stool, he smoked, his eye half closed. In a hoarse voice, he asked:

"Where are you going?"

"To take a walk. It's too hot."

"____6______--everything's closed. And no streetlights around here. You'd better stay put." I shrugged my shoulders, ____7______ "back soon," and plunged into the darkness. At first I couldn't see anything. I _____8_____ along the cobblestone street. I lit a cigarette. Suddenly the moon appeared from behind a black cloud, lighting a white wall that was crumbled in places. I stopped, blinded by such whiteness. Wind whistled slightly. I breathed the air of the tamarinds. The night ______9______, full of leaves and insects. Crickets ____10______ in the tall grass. I raised my head: up there the stars too had set up camp. I thought that the ____11______ was a vast system of signs, a conversation between giant beings. My actions, the cricket's saw, the star's blink, were nothing but pauses and syllables, ____12_____ phrases from that dialogue. What word could it be, of which I was only a syllable? Who speaks the word? To whom is it spoken? I threw my cigarette down on the sidewalk. Falling, it drew a shining curve, shooting out brief sparks like a tiny comet.

I walked a long time, slowly. I felt free, secure between the lips that were at that moment speaking me with such happiness. The night was a garden of eyes. As I crossed the street, I heard someone come out of a doorway. I turned around, but could not distinguish anything. I hurried on. A few moments later I heard the dull shuffle of ____13______ on the hot stone. I didn't want to turn around, although I felt the shadow getting closer with every step. I tried to run. I couldn't. Suddenly I stopped short. Before I could defend myself, I felt the point of a knife in my back and a sweet voice;

"Don't move, mister, or I'll stick it in."

Without turning, I asked;

"What do you want?"

"Your eyes, mister," answered the soft, almost painful voice.

"My eyes? What do you want with my eyes? Look, I've got some money. Not much, but it's something. I'll give you everything I have if you let me go. Don't kill me."

"Don't be afraid, mister, I won't kill you. I'm only going to take your eyes."

"But why do you want my eyes?" I asked again.

"My girlfriend has this ____14______. She wants a bouquet of blue eyes. And around here they're hard to find."

"My eyes won't help you. They're brown, not blue."

"Don't try to fool me, mister. I know very well that yours are blue."

"Don't take the eyes of a fellow man. I'll give you something else."

"Don't play saint with me," he said harshly. "Turn around."

I turned. He was small and fragile. His palm sombrero covered half his face. In his right hand he held a country machete that shone in the moonlight.

"Let me see your face."

He struck another match, and put it near my eyes. ____15______ my sleeve, he ordered:

"Kneel down."

I knelt. With one hand he grabbed me by the hair, pulling my head back. He bent over me, curious and tense, while his machete slowly dropped until it grazed my eyelids. I closed my eyes.

"Keep them open," he ordered.

I opened my eyes. The flame burned my lashes. All of a sudden he let me go.

"All right, they're not blue. Beat it." He vanished. I leaned against the wall, my head in my hands. I pulled myself together. Stumbling, falling, trying to get up again. I ran for an hour through the ____16______ town. When I got to the plaza, I saw the owner of the boardinghouse, still sitting in the front of the door. I went in without saying a word. The next day I left town.

ver. 3

Appendix 6 .2 : Data from Blue Bouquet Gap Filling

Bold words are word the researcher deemed as acceptable substitutes. Words with an asterisk following were translated from Japanese.

Targe Word	*N of Yes*	Y/N	Responses
sweat		N	satisfaction, eyelids, scorpions, women, my lover's arms, my lover, very fat cat, big leaf, cats and dogs, snow, dog's excrement, perplexity, pleasure, flower, white snow, blanket, sun shine
dazzle		N	agressive, flying, moth, flying like silk, a firefly, pretty girl, which entered from my room's window, strange eyes, took wing, beautifully, which was very beautiful, which was very beautiful, flying around flower, which is very big, and moths, honey bee
barefoot		N	at once, slow, across, narrow, slowly, all the day, that the steam-filled, dropped, to catch up with the butterfly, quickly, where was very ugly, where dust and junk were scattered in, slowly, slowly, without sounds
taciturn	6	Y	**piratical**, had, rough, people, catching, turned, called 'another eyed', **strange**, little eaten, and long haired, small, looked up to, **kind of scary**, panched or loosing, **terrible**, **poor**, **old**
shrug	3	Y	clean and beautiful, dark, **dropped**, put, pitch dark, lend, put my hand on, want to bring, got hurt, hit, **pushed**, lost, was hit, seized, **hung**, was patted on, hurt, hurt
cobblestone		N	clean, beautiful, dark, sesami, march, pitch-dark, curved, big, long, narrow, long, small, dark, quiet, sweet honey, old walls of the, wide, most dark, silent, old, foggy, short, lonely, big, very long, long, large
crumble	4	Y	hung, interested, mounted, beautiful, secret, **like a Dali painting (disastrous)**, happiness, shining strangely, big, the most high wall, **covered with thorns**, standing, the highest, **broken**, hiding
saw	5	Y	behaviour, bivouacking, rule, wings, the globe, dance, **chirp**, **small voice**, mind, show was popular in this place, dance, jumping, **crying**, **sounds**, **singing**, elegant playing,
sidewalk		N	table, street, main street, tall grass, ground, foot, stone, small pond near the street, dustbin, head,

			horizon, street, ground, pond, ground
comet	5	Y	stone, **lightning**, treasure box, **stars**, kangaroo, **jewellery**, peacock, **star**, thunder, afternoon, **star**, little boy, bird
distinguish	8	Y	**see**, find, do, feel, **see**, **see**, **see**, shout, say,turn, heard someone did, **find**, meet, **see**, **find out**, do, **see**,
bouquet	1	Y	husband, **bouquet**, jewel, piece, bright, girlfriend, look expensive, handsome gentleman, something, mystic charming, beauty, beautiful eyeball, child, became clear,
sombrero		N	perfectly, was, suddenly,was, peeped, firefly, small, old, was, that was ugly, doubtfully, finger, hands, was, wounded,
machete	4	Y	**stone**, **knife**, house, so, **knife**, map, spirit, appeared, lake, **stone**, pigeon blood, side, party, nostalgia
lashes	6	Y	heart, **eyes**, heart, head, **eyelash**, head, **eyes**, **eyes**, **eyes**, fear, **eyes**, clothes, afraid
stumble	2	Y	In fact, Then, Thinking, getting away, understanding, roll, **running** and shouting, carefully, flying, flying, floating, **rolling over stairs**, In addition, I was desperately, stupidly
boardinghouse	5	Y	**shop**, **motel**, chopsticks, back, government, lifestyle, **hotel at the strange town**, group, most bad man, **room**, blue eyes, **boardinghouse**
spray		N	black, brick, sounds, sounds, ceiling, sakura, human, dog, cat, garden, flower bed, sunshine, sun,
scorpion			sweat air, cat, butterfly, dogs, dog,
inhale	7	Y	**into**, **breathed in***, changed, **absorbed***, felt, **breathed in***, eat, **breath**, **breath**, **blow**
pewter			bath, wet, hand, bathroom
fold	2	Y	**pocket**, behind, **pocket***, middle, middle, box, left
wicker		Y	**plastic***, my, bench, chair
hoarse	3	Y	big, fear, small, very big, big, **small***, **small**, fox, dog, children, big, **low**, crazy, big
streetlights		N	take a walk, looking, house, everyone, food, wife, world, bird, man, people, water
plunge	2	Y	run, ran, around, sky, look, children, cat, **go**, run, **disappeared**, room, room, sky
tamarind	3	Y	**green**, night, with my mouth*, outside **freshness***, human, white, night, darkness, depth*, **country**, mountain,
blink	4	Y	me, **lightning**, very dark, cow, **shining**, **sparkle***, saw, **shine**

shuffle		N	left, red-coloured*, brother, pair, one, two, burning
saint		N	take, football, to, baseball, games, mother
palm		N	eyes, hand shadow, mother
graze		N	was in, sink
grey-winged		N	firefly, house, eyes, car, condition, blue, sea, beautiful, small, big, white, beautiful, wonderful, beautiful, small, small,
hammock	9	Y	hand, **bed**, room, **bed**, **bed**, face, **bed**, **bed**, **bed**, **bed**, **bed**, house, area, **bed**, foot, garden, standing place, house, room,
hideout	8	Y	**room**, foots, heart, **house**, hand, **room**, **room**, **house**, **house**, around, leg, **room**, **room**, heart, breath, hand, **room**,
enormous		N	moon, boared, gentleman, mouse, excite, go home, strange, moon, nature, scorption,
rub	8	Y	wished, **washed**, wet*, wet*, **stroked***,wet, **washed**, **wash**, **wiped***, sat, **wash**, have, wet*, wet, **wiped***, **wash**, put off, wear,
Hmmm	3	Y	**OK**, **now**, street, store, store, window, shops, jump, already, around, **and said**, called, arms,
mutter	5	Y	knees, **answered***, pushed, hands, **cry**, **said**, flying, listening, **and said**, **called**, arms
fumble		N	walk, came, walk, am not, run, through, through, walked, walk, run, was, walked, walked, go, walked,
hum	2	Y	playing, is shine, raised, silent, work, strong wind, cold, everyone hear **voice**, later, street, eyes, is , **surmise**, is mysterious.
bivouac	3	Y	play, are moving, stand, swimming, **are hiding***, play, were, **covered**, are, spend, **hid**.
universe	3	Y	computer, stone, heart, eat, nothing, signal, **the stars**, **star**,
scatter		N	mathematics, economy, sun, nothing, eat, nothing, it, board, and
sandals		N	pan, something*, someon,water, voice, cooking, nothing, eyes, my, lips, card,
whim		N	watch, beautiful eyes, brown eyes, eyes, eyes, sick, eyes, same eye, egg, brown eyes, same eyes, eyes, black eyes, one,
grab	1	Y	please, he **pat/touch***, cutting, suddenly,
deserted	2	Y	down, **ghost**, left, left, **strange**, dish, down,